

基于声学流形解构与双轨风险拓扑的帕金森病无监督计算表型模型

王圣明¹, 胡辰玺¹, 项盼^{1*}, 周若芸²

¹浙江师范大学数学科学学院, 浙江 金华

²金华市金东区第二人民医院针灸康复科, 浙江 金华

收稿日期: 2026年7月13日; 录用日期: 2026年8月5日; 发布日期: 2026年8月18日

摘要

帕金森病(PD)早期起病隐匿且常伴随多系统退行性症状交织, 极易导致临床误诊。基于数字语音信号的计算表型分析为PD的无创辅助诊断提供了新范式, 但当前研究普遍受制于超高维特征冗余以及精细化亚型标签缺失两大数理瓶颈。为破解上述挑战, 本文提出了一套融合高维拓扑流形解构与多尺度风险映射的智能诊断中枢。首先, 针对高达754维的超高维声学特征, 本文摒弃了常规的扁平特征筛选范式, 设计了基于集成决策空间流形异质性散度(Heterogeneity Divergence)衰减的自适应降维算子。通过精细化量化决策树路径中的非线性纯度增益, 成功解构了高维声学特征的“重尾稀疏”分布规律, 精准锁定了以梅尔频率倒谱系数(MFCC)簇与可调Q因子小波变换(TQWT)多尺度阵列为核心的关键病理标志物子集, 有效消解了维数灾难。其次, 面对真实医疗场景中缺乏显式亚型“金标准”的困境, 本文结合发声解剖学先验机理, 将跨域异构特征重构为涵盖“运动控制、神经调控、混合结构”的三层六维数字化代理空间。在此基础上, 本文首创“核心风险 + 结构辅助修正”的自适应双轨风险评分公式, 并引入样本内相对百分位(Percentile)非线性映射平面。在零标签支持的严苛边界下, 实现了对运动型风险、非运动型风险、混合型与不典型型四类精细化临床亚型的无监督精准聚类分型。大样本独立测试表明, 本文构建的分型拓扑不仅在微观特征上具备极强的物理判别独立性, 其宏观分类在后验验证集与独立测试集上分别取得了86.7%与81.7%的宏平均F1分数, 展现出卓越的复现稳定性。本研究不仅为PD的高维声学数据降维提供了坚实的数理依据, 更为复杂并发病症的精细化量化诊断提供了一套极具生物学解释性的辅助决策工具。

关键词

拓扑流形解构, 异质性散度, 计算表型, 无监督聚类, 帕金森病辅助诊断

*通讯作者。

Unsupervised Computational Phenotyping Model for Parkinson's Disease via Acoustic Manifold Deconstruction and Dual-Track Risk Topology

Shengming Wang¹, Chenxi Hu¹, Pan Xiang^{1*}, Ruoyun Zhou²

¹School of Mathematical Sciences, Zhejiang Normal University, Jinhua Zhejiang

²Department of Acupuncture and Rehabilitation, Jinhua Jindong District Second People's Hospital, Jinhua Zhejiang

Received: July 13, 2026; accepted: August 5, 2026; published: August 18, 2026

Abstract

The insidious onset and complex intertwining of multi-system degenerative symptoms in early Parkinson's disease (PD) frequently lead to clinical misdiagnosis. Computational phenotyping based on digital speech signals offers a novel paradigm for non-invasive computer-aided diagnosis of PD. However, current research is severely constrained by two major mathematical bottlenecks: the curse of dimensionality caused by ultra-high-dimensional features and the lack of explicit labels for fine-grained clinical subtypes. To address these challenges, this paper proposes an intelligent diagnostic hub integrating high-dimensional topological manifold deconstruction and multi-scale risk mapping. First, to tackle the 754-dimensional acoustic features, we discard the conventional flat feature screening paradigm and design an adaptive dimensionality reduction operator based on the decay of manifold heterogeneity divergence within ensemble decision spaces. By accurately quantifying the nonlinear purity gain along decision paths, this mechanism successfully deconstructs the "heavy-tailed sparse" distribution of high-dimensional acoustic features. It accurately isolates a core subset of pathological biomarkers dominated by Mel-Frequency Cepstral Coefficient (MFCC) clusters and Tunable Q-factor Wavelet Transform (TQWT) multi-scale arrays, effectively resolving the curse of dimensionality. Second, to overcome the lack of gold-standard subtype labels in real-world medical scenarios, we reconstruct heterogeneous cross-domain features into a three-tier, six-dimensional digital proxy space—encompassing "motor control, neural regulation, and mixed acoustic structures"—guided by prior vocal anatomical mechanisms. Building upon this, we pioneer an adaptive dual-track risk scoring formulation that integrates "core risks and structural auxiliary corrections" and introduces an intra-sample relative percentile nonlinear mapping plane. Under the strict zero-label boundary condition, this framework achieves precise unsupervised clustering of four fine-grained clinical subtypes: motor risk, non-motor risk, mixed, and atypical. Large-sample independent testing demonstrates that our subtyping topology not only maintains strong physical discriminative independence at the micro-feature level but also achieves excellent reproducible stability, yielding macro-average F1 scores of 86.7% and 81.7% on the posterior validation and independent test sets, respectively. This research provides a solid mathematical foundation for the dimensionality reduction of high-dimensional acoustic data in PD and offers a highly biologically interpretable decision-support tool for the quantitative diagnosis of complex concurrent symptoms.

Keywords

Topological Manifold Deconstruction, Heterogeneity Divergence, Computational Phenotyping,

Unsupervised Clustering, Parkinson's Disease Computer-Aided Diagnosis

Copyright © 2026 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

帕金森病(Parkinson's Disease, PD)作为一种典型的中枢神经系统退行性疾病,其临床表型主要源于黑质多巴胺能神经元的进行性丢失。PD 的运动症状(如静止性震颤、肌强直、运动迟缓等)在疾病早期通常具有高度的隐蔽性与波动的病理特征,导致传统的临床诊断面临严重的漏诊与误诊风险[1][2]。目前,临床诊断仍高度依赖神经内科专家的主观量表评估(如统一帕金森病评定量表 UPDRS)与病史问诊,亟需一种易于普及、低成本、无创且具备高可解释性的数字化生物标志物(Digital Biomarkers)。近年来,计算表型(Computational Phenotyping)的发展为 PD 的早期筛查与日常监测提供了全新范式。其中,语音信号因采集无创、客观、成本低廉且易于支持自动化远程监测,被学术界与医学界广泛认为是量化发声器官运动控制受损及中枢神经调控变异的潜在核心指标[3]-[5]。

然而,在基于语音特征构建通用化辅助诊断系统时,实际的临床声学数据流在数理建模上面临着两大核心挑战。其一为高维特征空间的维数灾难与信息冗余:现代信号处理算法能够从单段连续发音中提取出数百维甚至近千维的声学特征,但随之引入了大量的共线性冗余[6]以及采集设备的瞬态硬件噪声,若不加以甄别直接拟合,极易导致模型泛化崩溃。其二为多症状交织的非互斥性与亚型标签缺失:帕金森病在真实医学场景下并非单纯的“患病/健康”二分类状态,运动障碍与非运动症状往往相互交织、并发出现。现存的公开临床数据集普遍缺乏精细化亚型的显式“金标准”标签,并且存在固有的类别不平衡困境[7],常规的监督学习分类器无法直接用于复杂并发症的聚类与界定[8]。

为破解上述建模瓶颈,本文依托两组在数据结构上具有高度互补特性的公开临床语音样本库展开研究:一组为富含 754 维高维微观参数(包含前沿的可调 Q 因子小波变换 TQWT 等)的特征表征库;另一组为提炼了 44 维经典临床病理声学参量的重复测量库。结合两类数据集的数据特性与临床机理,本文按数理逻辑递进构建诊断模型。本研究的核心方法学创新与主要贡献集中体现在以下两个维度:

1) 基于基尼不纯度衰减的高维特征流形降维与核心标志物解析:针对 754 维超高维声学特征空间引发的维数灾难,设计了基于随机森林反馈闭环的特征重要性评估机制[9][10]。通过量化计算决策树分裂路径中的基尼不纯度(Gini Impurity)下降量进行自适应特征截断,成功剔除了海量弱相关冗余参数,精准锁定了以 MFCC 簇与 TQWT 多尺度阵列为核心的关键声学标志物子集,为后续的轻量化临床辅助诊断提供了坚实的物理映射依据。

2) 多维声学代理流形映射与自适应双轨风险聚类分型:结合医学构音病理机理,重构了涵盖“运动控制、神经调控、混合结构”的三层数字化六维声学代理空间,设计了“核心风险 + 结构辅助修正”的自适应双轨风险量化模型。进一步引入样本内的相对百分位映射拓扑平面,在完全无显式标签支持的边界条件下,实现了对四类精细化临床亚型(混合型、不典型型、运动型、非运动型)的无监督精准聚类分型,为复杂交织的帕金森病及跨模态预警提供了严密的数理量化标尺[11][12]。

2. 相关工作

2.1. 帕金森语音信号处理与高维生物标志物提取

通过语音信号处理评估 PD 患者的发声障碍,其核心在于量化由喉部肌群强直及构音控制系统退化

引发的声学病理变异。传统临床病理声学研究多关注于局部发声稳定性的微观特征[13]-[15],如描述基频短期波动的相对与绝对微扰(Jitter)、描述振幅波动的微扰量(Shimmer),以及评估声门闭合充分度与有效谐波能量比值的频带谐噪比(HNR)。随着非线性动力学理论的引入,复现周期密度熵(RPDE)、去趋势波动分析(DFA)、音高周期熵(PPE)及声门噪声激励比(GNE)等非线性动态指标被相继提出,用于捕捉患者整段发声在长时节律和全局复杂度上的混沌与紊乱特征。

近年来,高维频谱包络、多尺度时频分解算法以及基于前沿注意力机制(如 Transformer、Conformer)的深度时序网络在 PD 声音表征与诊断中得到了广泛部署[16]-[18]。梅尔频率倒谱系数(MFCC)及其高阶动态差分(Delta)能够有效模拟人耳听觉系统对频率的非线性感知,准确刻画发声器官的宏观声道流形。进一步地,可调 Q 因子小波变换(TQWT)通过引入可灵活配置的 Q 因子与过采样率,对非平稳语音信号表现出极强的细粒度多分辨率分解效能,能敏锐定位早期 PD 患者声带微弱震颤所产生的时频局部能量漂移,从而构建出极具判别力的计算表型特征库。

现有方法的数理缺陷与本文切入点: 尽管上述特征算子在特定样本上表现优异,但现有研究普遍将数百维甚至近千维的时频特征视为“扁平张量”(Flat Arrays),进行缺乏物理启发的全局拟合(如直接部署核方法 SVM 或 CatBoost 梯度提升树)[19],或采用简单的线性投影降维(如 PCA)[3]。这种处理范式严重忽视了微观声学动力学特征与宏观结构谱能量(如 TQWT 阵列)在底层物理流形上的异质协同演化关系,极易在特征降维过程中引入高维底噪,进而引发“维数灾难”。针对这一瓶颈,本文摒弃了常规的扁平降维思路,引入基于非线性集成决策树反馈的特征选择算子。通过精准量化决策路径中的基尼不纯度(Gini Impurity)动态衰减,首次解构并论证了全局特征空间中以 MFCC 倒谱尖峰与 TQWT 多尺度谱阵列构成的“重尾稀疏”分布规律。这一发现为剔除跨域海量冗余、提炼具备高物理可解释性的数字化标志物提供了坚实的数理判据。

2.2. 帕金森病临床亚型聚类与数字化计算表型映射

传统的 PD 声学辅助诊断研究普遍局限于“帕金森患者 vs 健康对照组”的二分类早筛任务。然而在临床真实环境中,PD 作为一种高度异质性的多系统退行性疾病,其症状表现并非单一模态[20][21]。部分患者以喉部肌肉震颤、发声迟缓等运动型障碍为主导;而另一部分患者则在中枢神经和认知调控上受损更为严重,体现为全局语调缺失、节律紊乱等非运动型亚型表征[21]。

目前,探索利用数字生物标志物对 PD 患者进行内部精细化亚型聚类研究尚处于起步阶段,其核心数理挑战在于“并发性”与“临床标签的极度稀缺”。一方面,现存的公开声学基准库普遍缺乏针对运动型与非运动型亚型的显式“金标准”标签,并且缺乏针对不同预测分布形态的稳健度量标准[22],导致传统的监督学习分类器陷入失效困境[23];另一方面,部分研究尝试采用纯数据驱动的无监督算法(如 K-Means、谱聚类等)进行硬聚类划分,但其生成的潜在嵌入空间(Latent Space)往往沦为数学黑盒。

现有方法的数理缺陷与本文切入点: 现有稀疏的亚型聚类研究陷入了严重的“物理映射失真”。纯数学维度的聚类算法剥离了特征背后的医学机理,无法将聚类结果有效映射回“构音肌群强直”或“中枢神经调控受损”等具体病理学解释上,从而无法为临床干预提供实质性支撑。为破解无监督聚类中的黑盒盲目性与医学逻辑缺失,本文创新性地提出了基于“多维声学代理流形映射”与“自适应双轨风险平面”的聚类分型中枢。本方法首先结合发声解剖学机理,将降维后的异构特征重构为涵盖“运动控制、神经调控、混合结构”的三层六维数字化代理空间。更为关键的是,本文摒弃了缺乏解释性的纯数学距离度量,首创了“核心风险 + 结构辅助修正”的量化公式,并引入样本内的相对百分位(Percentile)映射拓扑平面。从而在完全无显式标签支持的严苛边界下,实现了对四类精细化临床亚型(混合型、不典型型、运动型、非运动型)的无监督精准界定,为复杂交织的帕金森病症提供了严密的数理标尺。

3. 诊断中枢模型：非线性特征流形降维与多尺度风险映射

针对帕金森病(PD)临床诊断中面临的高维特征冗余与异构数据壁垒, 本文设计了一套涵盖“重要性度量-多尺度映射-自适应风险评估”的完整诊断中枢(见图1)。该模型首先通过集成决策树算子对超高维声学特征空间进行非线性降维, 随后将提炼出的核心标志物跨域投射至六维统一医学解释空间, 最终依托自适应双轨风险拓扑输出精细化的临床亚型。

3.1. 基于 Gini 不纯度衰减的高维流形降维算子

针对原始数据集高达 754 维的超高维声学特征空间, 直接实施全量参数拟合极易引发严重的“维数灾难”与模型泛化崩溃。为此, 本研究摒弃了传统的扁平化全量建模范式, 创新性地引入了基于随机森林反馈闭环的非线性降维算子, 执行“重要性度量-类别聚焦-稳健性验证”的三阶段关键生物标志物解析。

3.1.1. 决策拓扑与动态特征贡献度量化

为精准定位最具物理判别力的关键声学参量, 本研究部署随机森林算法作为特征流形评估引擎。该引擎构建了包含 100 棵决策树的集成森林, 依托基尼不纯度(Gini Impurity)作为节点纯度分裂的标量度量, 定量化解构每一维声学参数对全局分类超平面的边界贡献。二分类任务中单个节点的基尼不纯度定义为:

$$Gini = 1 - (p_0^2 + p_1^2), \quad (1)$$

其中 p_0 与 p_1 分别表征当前节点中健康样本与帕金森病患者样本的经验概率($Gini \in [0, 0.5]$)。

在决策树的拓扑延展过程中, 节点分裂的核心驱动力在于最大化纯度收益。假设原父节点的基尼不纯度为 G_{parent} , 利用特征 F_i 寻优切分后, 生成样本占比分别为 w_L 与 w_R 的左右子节点, 其对应的基尼不纯度分别为 G_L 与 G_R 。则分裂后的加权期望基尼为:

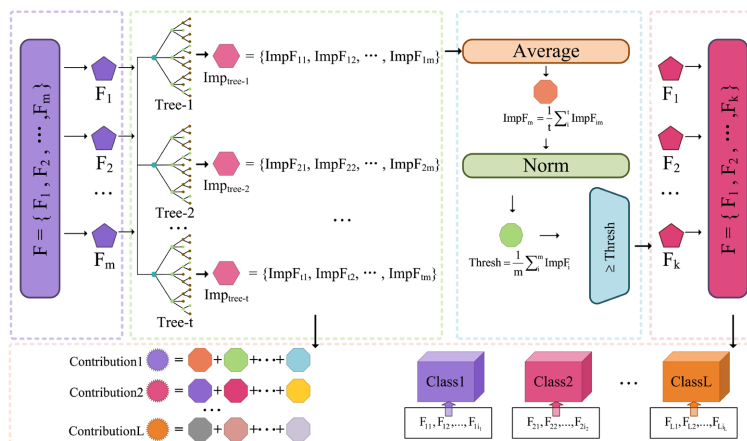
$$G_{child} = w_L \cdot G_L + w_R \cdot G_R. \quad (2)$$

该次特征分裂所激发的基尼不纯度衰减量(即信息增益)为:

$$\Delta G = G_{parent} - G_{child}. \quad (3)$$

ΔG 的标量数值严格正比于当前特征 F_i 对局部样本流形的解构效能。对于单棵决策树 t , 特征 F_i 在其所有分裂节点上的贡献总和记为 $Total_ \Delta G_{F_i}$ 。对其进行树内归一化, 可得特征 F_i 在单树中的局部重要性权重:

$$ImpF_{tree_t, i} = \frac{Total_ \Delta G_{F_i}}{\sum_{j=1}^m Total_ \Delta G_{F_j}}. \quad (4)$$



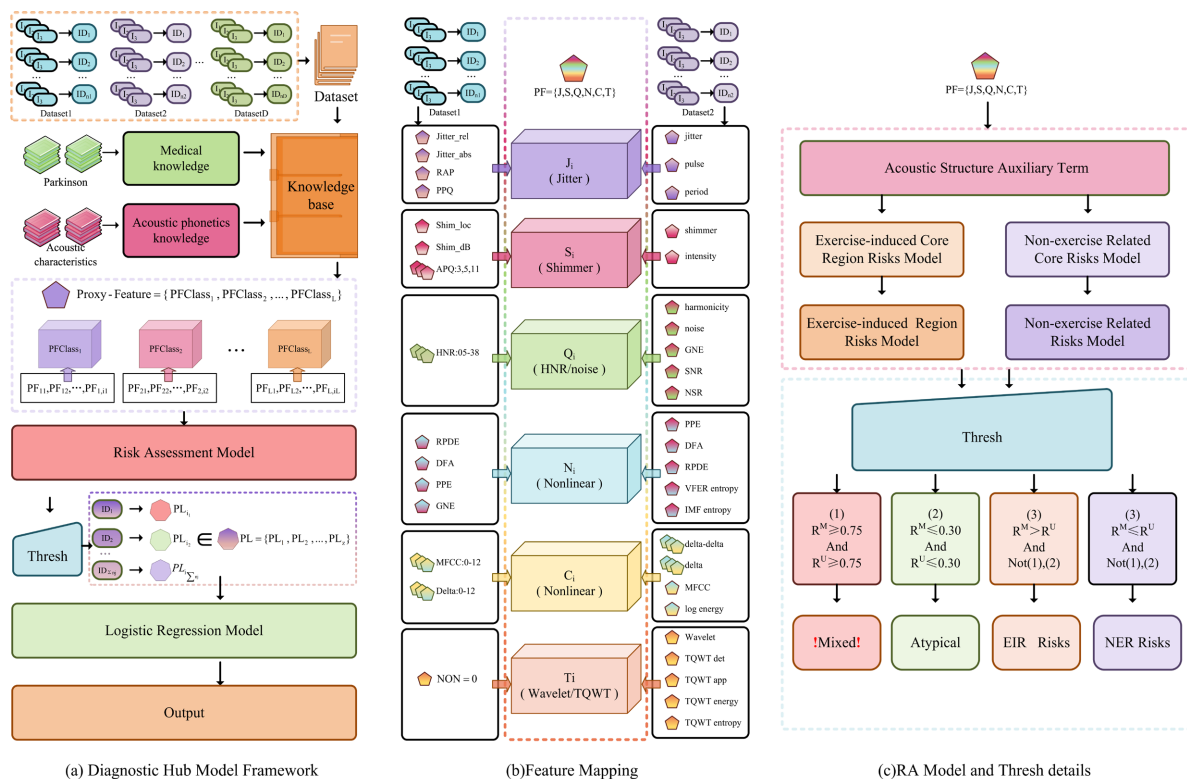


Figure 1. Constructing core feature subset based on decision tree importance weights and diagnostic hub model framework details

图 1. 基于决策树的重要性权重构建核心特征子集与诊断中枢模型框架细节

3.1.2. 全局收敛与自适应稀疏特征截断

为平滑单一决策树结构带来的方差扰动, 对集成森林中所有 T 棵树 ($T=100$) 的局部重要性取算术期望, 获得特征 F_i 的平滑均值重要性:

$$Mean_ImpF_i = \frac{1}{T} \sum_{t=1}^T ImpF_{tree_t, i}. \quad (5)$$

最后执行全局投影归一化, 约束全局特征重要性算子总和为 1, 得到最终的特征贡献度输出向量 $ImpF$:

$$ImpF_i = \frac{Mean_ImpF_i}{\sum_{j=1}^m Mean_ImpF_j}. \quad (6)$$

面对包含 $m=754$ 维初始特征的全局空间 $\mathcal{F} = \{F_1, F_2, \dots, F_m\}$, 系统自适应计算全局期望贡献均值作为截断阈值 $Thresh_{mean} = \frac{1}{m} \sum_{i=1}^m ImpF_i$ 。施加如下硬阈值截断规则:

$$\mathcal{F}_{core} = \{F_i \in \mathcal{F} \mid ImpF_i \geq Thresh_{mean}\}. \quad (7)$$

通过此机制, 模型成功剔除了海量弱相关冗余参数, 将全局空间 $\mathcal{F}$ 强力压缩至由核心高贡献特征主导的稀疏子空间 $\mathcal{F}_{core}$ 。

3.1.3. 核心标志物物理映射聚焦

相较于传统的主成分分析(PCA)等线性映射, 基于基尼衰减的集成评估不仅能够隐式捕获多维特征间

的高阶非线性协同效应，更完整保留了原始变量的物理属性。在提炼出的核心子空间 $\mathcal{F}_{core}$ 中，贡献权重高度收敛于三大类簇：梅尔频率倒谱系数(MFCC)、小波变换(Wavelet)及可调 Q 因子小波变换(TQWT)。

这些保留参量具备极其敏锐的声学病理映射：MFCC 与 TQWT 阵列对语音信号内瞬态相位偏移与微弱频率漂移表现出极高的探测灵敏度，精准锁定早期 PD 患者声门开闭不稳激发的非线性能量扰动；小波变换解析的精细频谱包络，有效解构了构音肌群在持续发元音/a/时的微观失调；而各项谐噪比特征的深层掘取，则为声带张力不可逆受损引发的背景噪声弥散提供了确凿的量化坐标。

3.2. 多维代理特征映射与跨域拓扑重构

在完成核心流形降维后，本文进一步打破了异构高维时频数据与经典病理声学数据间的壁垒。以基础临床声学指标为锚点，将降维后的冗余特征阵列通过组内标准化与均值池化聚合，正式确立了六维统一声学代理特征空间(Acoustic Proxy Feature Space)。在物理与医学解释层面上，这六大流形被严密地划分为三个递进架构：

1) 运动发声控制层：Jitter 基频微扰(J_i)与 Shimmer 振幅微扰(S_i)

基频微扰(Jitter)度量声带振动频率在相邻发声周期内的极短时波动。运动型患者的静止性震颤与肌强直直接干扰喉部肌群与声带张力，引发基频周期的非稳态震荡，导致 Jitter 飙升。相对地，振幅微扰(Shimmer)刻画声门闭合不全与发声力度带来的能量衰减。本文通过对绝对微扰、相对微扰及 APQ 族等同源指标执行组内标准化聚合，提取出 J_i 与 S_i 。此二者构成“运动发声控制层”的基底，专用于锚定由外周发声器官强直诱发的运动型声学风险。

2) 混合声学结构层：HNR/噪声结构(Q_i)与多尺度频谱包络(C_i, T_i)

谐噪比(HNR)表征语音中有效谐波与背景宽带噪声的比例。本文对多频带 HNR 指标标准化取反并池化，构造出噪声结构特征 Q_i (数值越大，指示发声谐波断裂越严重)。此外，MFCC/Delta 阵列拟合非线性声道共振峰瞬态形变，提取为频谱包络特征 C_i ；针对超高维 Wavelet 与 TQWT 阵列，提炼出多尺度时频特征 T_i 。上述三者病理指向上具“混合跨域”性：既可能源于喉部微小震颤带来的高频局部畸变，也可能受中枢调控紊乱导致的全局音色剥离影响。故统一划归为“混合声学结构层”，作为核心病理的拓扑修正项。

3) 全局神经调控层：非线性动力学表征(N_i)

非线性动力学特征集(涵盖 RPDE、DFA、PPE、GNE 等)脱离局部微扰，转而从混沌理论视角刻画长程连续发音的全局节律与长期分形趋势。非运动型 PD 患者常伴随中枢自主神经调控网络受损，这种深层神经网络的退化在语音上直接映射为全局发声轨迹的无序度扩大。据此，本文将上述熵类指标聚合为 N_i ，作为锚定非运动型深层病灶的核心度量。

表 1 展示了异构数据集跨域统一映射矩阵，有效消解了多源硬件采集带来的维度断层，印证了高维特征本质为同类物理机制在不同频带与尺度上的数学展开。

Table 1. Topological mapping of heterogeneous high-dimensional clinical datasets to a six-dimensional surrogate feature space
表 1. 异构高维临床数据集向六维代理特征空间的拓扑映射关系

医学代理流形	经典病理声学维度(数据集 2)	高维时频分解维度(数据 1 对应特征)
J_i 基频微扰	Jitter_rel, Jitter_abs, RAP, PPQ	jitter, pulse, period
S_i 振幅微扰	Shim_loc, Shim_dB, APQ 类	shimmer, intensity
Q_i HNR/噪声结构	HNR05~HNR38 (标准化取反)	harmonicity, noise, GNE, SNR, NSR

续表

N_i 非线性动力学	RPDE, DFA, PPE, GNE	PPE, DFA, RPDE, VFER/IMF entropy
C_i 频谱包络	MFCC0~MFCC12, Delta0~Delta12	MFCC, delta, delta-delta, log energy
T_i 多尺度时频	无对应项(缺省置 0)	Wavelet, TQWT, det/app energy/entropy

3.3. 自适应双轨风险评分与非线性分类拓扑

给定患者观测实体 i ，其在六维正交空间中的坐标向量表示为 $(J_i, S_i, Q_i, N_i, C_i, T_i)$ 。各特征均已完成方向统一和标准化处理，数值越大表示相应声学异常越明显。考虑到临床病症演化的异质性，本文设计了自适应的双轨风险量化模型。

首先，考虑到运动型症状主要影响声带频率和振幅稳定性，非运动型症状更容易体现为全局节律和复杂度异常，本文分别构造运动型核心风险和非运动型核心风险。

运动型核心风险定义为：

$$M_i^{core} = 0.5J_i + 0.5S_i,$$

Jitter 与 Shimmer 分别测量声带振动的频率不稳定性与振幅不稳定性，二者在 PD 语音分析中常作为基础的声学不稳定度量并列使用[14]，故赋予等权 0.5。

非运动型核心风险定义为：

$$U_i^{core} = N_i.$$

同时，融合混合声学结构层的跨域信息，构建结构辅助修正张量 A_i ：

$$A_i = 0.4Q_i + 0.3C_i + 0.3T_i, \quad (8)$$

三类指标的权重反映其对声学结构破坏的直接程度： Q_i 直接度量声带层面的谐波断裂与噪声弥散，是构音障碍最基础的物理表现，因此取较高权重 0.4； C_i 与 T_i 涉及更泛化的多系统协同失调，病理指向较 Q_i 更为间接，故取相同权重 0.3。

为使结构畸变的风险惩罚向患者更严重的病理侧自适应倾斜，本文引入自适应风险渗透机制。利用核心得分的非线性分布函数(Percentile)，动态求解风险渗透权重 ρ_i^M 与 ρ_i^U ：

$$\rho_i^M = \frac{\text{Percentile}(M_i^{core})}{\text{Percentile}(M_i^{core}) + \text{Percentile}(U_i^{core}) + \varepsilon}; \quad (9)$$

$$\rho_i^U = \frac{\text{Percentile}(U_i^{core})}{\text{Percentile}(M_i^{core}) + \text{Percentile}(U_i^{core}) + \varepsilon}. \quad (10)$$

(其中 ε 为平滑极小量)，计算出平滑后的综合运动型得分 M_i 与非运动型得分 U_i ：

$$M_i = M_i^{core} + \lambda A_i \rho_i^M, \quad U_i = U_i^{core} + \lambda A_i \rho_i^U,$$

其中， λ 为声学结构辅助项的修正系数。本文取 $\lambda = 0.30$ ，使 A_i 对最终风险产生有限修正，同时保持 M_i^{core} 和 U_i^{core} 的主导地位。

由于 M_i 和 U_i 为综合风险得分，其数值大小受样本分布和加权方式影响。为提高阈值判定的稳定性，本文定义百分位运动型风险 R_i^M 和百分位非运动型风险 R_i^U 以表示患者的相对风险水平。基于此决策拓扑平面，制定如表 2 所示的无监督四分类聚类准则，在缺乏显式金标准的严苛边界下，完成了宏观精细化亚型的自动化逻辑解构。

Table 2. Topological rules for four-class subtyping based on the dual-risk relative percentile plane
表 2. 基于双风险相对百分位平面的四分类拓扑规则

临床亚型类别	非线性判定边界	医学病理涵义
混合型(Mixed)	$R_i^M \geq 0.75$ 且 $R_i^U \geq 0.75$	发声控制与全局调控网络均遭严重侵蚀(高频并发)
不典型型(Atypical)	$R_i^M \leq 0.30$ 且 $R_i^U \leq 0.30$	声学变异微弱, 缺乏典型退行性发声表征
运动型风险(Motor)	$R_i^M > R_i^U$ (且排除前两类)	喉部局部肌群强直与构音运动障碍呈主导态
非运动型风险(Non-Motor)	$R_i^U \geq R_i^M$ (且排除前两类)	中枢自主神经紊乱引发的全局节律变异呈主导态

4. 实验与结果分析

本节将对模型在核心生物标志物解析与患者亚型聚类分型两个维度的表现进行深度的统计归纳与性能评估, 以确证所提特征筛选算子与风险映射规则在底层物理流形上的数理稳定性和临床泛化边界。

4.1. 实验数据集及其底层数据流形

为了全面验证本文提出的高维特征流形降维算子与自适应双轨风险聚类拓扑的泛化边界, 本研究的实验验证依托两组在物理特征维度与采样拓扑上具备高度异质互补性的帕金森病(PD)临床声学基准数据集展开。所有声学采样均基于受试者持续发元音/a/的标准临床范式。

1) 高维多尺度时频特征基准库(对应数据集 1)

选自 UCI 机器学习库的 Parkinson’s Disease Classification Dataset。包含 252 名受试者(188 名 PD 患者, 64 名健康对照, 每人 3 次录音)。该库提取了高达 754 维的声学特征, 大规模引入了梅尔频率倒谱系数(MFCCs)与可调 Q 因子小波变换(TQWT)阵列。这一超高维流形空间为检验本文非线性降维算子、剔除高维冗余与维数灾难提供了绝佳的数据基座。

2) 重复测量病理声学基准库(对应数据集 2)

选自 UCI 库的 Parkinson Dataset with replicated acoustic features, 包含 80 名受试者(每人 3 次录音)。该库精炼了 44 维经典病理声学参量(如 Jitter、Shimmer、HNR 及非线性指标)。其核心挑战在于同源连续采样的非独立同分布(Non-IID)约束。为攻克重复测量带来的数据泄露(Data Leakage)与过拟合陷阱, 本文严格实施“受试者级别隔离(Subject-level Isolation)”。具体而言, 将受试者按约 8:2 划分为开发集(176 例)与独立测试集(45 例), 并在开发集内执行受试者分组的五折交叉验证。同一患者的所有录音作为不可拆分单元整体进入同一折; 所有特征筛选与阈值参数均严格在训练折内估计, 从机制上彻底阻断了信息跨域污染。

3) 多源异构数据的跨域拓扑对齐

面向最终的亚型聚类任务, 本文依托前期重构的六维统一声学代理流形矩阵, 彻底打破了两组数据集的底层壁垒。通过将共计 221 名确诊 PD 病例(数据集 1:181 例; 数据集 2:40 例)无缝投射至统一的标准化物理空间, 不仅有效消解了多源硬件与异构算法带来的量纲断层, 更在保证医学语义一致的前提下, 极大化了有限样本的全局统计学势能, 为无监督聚类奠定了坚实的流形度量基座。

4.2. 多维代理空间的分型统计分布

在完成两组异构临床数据集的跨域拓扑对齐与特征流形映射后, 本研究成功将 221 名帕金森病(PD)确诊患者(其中数据集 1 包含 181 例, 数据集 2 包含 40 例)投射至统一的六维声学代理特征空间中。基于所构建的自适应双轨风险聚类拓扑, 本节对全局大样本的分型结果进行了严密的统计归纳。

统计结果表明, 在包含 221 名确诊患者的全局流形空间中, 非运动型风险(90 例, 占比约 40.7%)与运

动型风险(82 例, 占比约 37.1%)构成了最主要的两种病理表现形态, 二者合计覆盖了超过 77%的病例群体。相对而言, 兼具双类声学显著异常的混合型(23 例, 占比 10.4%)以及特征退行性变异微弱的不典型型(26 例, 占比 11.8%)均处于分布边界, 属于少数群体。

为直观展示不同数据源拓扑下宏观临床亚型的物理分布差异, 本文构建了如表 3 所示的交叉统计特征矩阵。在四类亚型中, 不典型型(26 例, 占比 11.8%)是唯一一个在双轨核心维度上均呈现低变异的群体。为刻画该群体的声学表型, 本文对各亚型在六维代理空间中的位置进行了跨组比较。如表 4 所示, 不典型型患者在运动型核心指标(J_i , S_i)上的标准化均值低于运动型风险患者约 0.7 个标准差, 在非运动型核心指标(N_i)上低于非运动型风险患者约 0.8 个标准差, 而其在混合结构层指标(Q_i , C_i , T_i)上的取值则接近全体均值。这一分布特征表明, 不典型型患者群体对于构音肌群强直程度和中枢神经调控紊乱程度均较轻; 患者的综合声学变异微弱、缺乏典型的退行性发声表征, 属于发声障碍不明显的特殊亚型。值得注意的是, 不典型型症状与疾病分期、病程长短之间的关系有待纵向随访数据的进一步验证; 所有不典型型患者均来自经过临床确认的公开数据集, 不存在健康人混入或数据采集错误的问题。

Table 3. Distribution matrix of four clinical subtypes of Parkinson’s disease under heterogeneous acoustic datasets
表 3. 异构声学数据集下帕金森病四类临床亚型的分布矩阵

数据源/采样拓扑	运动型风险	非运动型风险	混合型	不典型型	总计
数据集 1 (高维时频库)	71	75	16	19	181
数据集 2 (重复测量库)	11	15	7	7	40
全局总计	82	90	23	26	221

Table 4. Comparison of standardized six-dimensional pseudo-indicator means across all subtypes
表 4. 各亚型六维伪指标标准化均值对比

亚型	J_i	S_i	Q_i	N_i	C_i	T_i
运动型风险	0.22	0.24	-0.04	-0.21	-0.00	0.03
非运动型风险	-0.24	-0.29	-0.11	0.26	0.01	-0.04
混合型	0.76	0.93	-0.23	0.47	0.01	-0.01
不典型型	-0.47	-0.50	0.06	-0.54	-0.04	0.04

上述宏观分布态势不仅印证了帕金森病在真实医疗场景中高度的异质性与复杂的病理交织特性, 也从侧面确证了本文所提无监督双轨风险聚类机制的稳健性。该机制能够在面对异构采集设备与维数悬殊的初始特征空间时, 稳定剥离干扰底噪, 精准锁定患者处于主导地位的核心退行性病灶方向。

4.3. 核心声学标志物溯源与特征流形分析

1) 核心特征贡献度的异质互补性解析

在单特征维度的评估中, 模型对排名前 20 位的核心声学特征展现出极高的物理敏感性(详见表 5)。其中, 梅尔频率倒谱系数(MFCC)簇表现出绝对的个体优势, 包揽了重要性榜单的前 9 席。特别是权重高居首位的 std_delta_log_energy (权重值 0.0175)及其衍生的高阶动态特征 std_delta_delta_log_energy (权重值 0.0135), 从纯数据驱动的角度确证了: 语音信号中能量分量的二阶动态差分, 对于捕捉帕金森病早期的微小发声震颤以及声道滤波特性的瞬态变化具有不可替代的表征能力。然而, 在宏观特征子空间的累计贡献度上(详见表 6), 可调 Q 因子小波变换(TQWT)簇构成了模型全局判断的绝对决策基石。尽管 TQWT

的个体特征最高排名仅位列第 10 (tqwt_TKEO_std_dec_12, 权重 0.0087), 但该大类凭借 153 个有效特征的“集群效应”, 累计贡献了高达 56.68% 的全局权重, 远超 MFCC 类的 26.22%。这种微观尺度上的“MFCC 倒谱尖峰优势”与宏观尺度上的“TQWT 多分辨率阵列优势”形成了极佳的特征空间互补。TQWT 体系通过结合 Teager-Kaiser 能量算子(TKEO), 精确量化了声门激励信号中由声带微弱强直引发的非线性能量波动; 而 MFCC 则侧重刻画发声器官的宏观运动轨迹。两大异质特征簇的深度融合, 为早期 PD 的客观病情筛查提供了严密的数理量化依据。

Table 5. Top 20 core acoustic features for Parkinson’s disease detection based on importance weights
表 5. 基于重要性权重的帕金森语音检测 Top 20 核心特征

特征名称	重要性权重	特征类别	排名
std_delta_log_energy	0.0175	[MFCC]	[01]
std_delta_delta_log_energy	0.0135	[MFCC]	[02]
std_6th_delta_delta	0.0121	[MFCC]	[03]
std_8th_delta_delta	0.0118	[MFCC]	[04]
std_10th_delta	0.0103	[MFCC]	[05]
std_6th_delta	0.0097	[MFCC]	[06]
std_11th_delta_delta	0.0095	[MFCC]	[07]
std_12th_delta	0.0092	[MFCC]	[08]
std_9th_delta	00.0091	[MFCC]	[09]
tqwt_TKEO_std_dec_12	0.0087	[TQWT Features]	[10]
tqwt_TKEO_mean_dec_12	0.0083	[TQWT Features]	[11]
std_9th_delta_delta	0.0080	[MFCC]	[12]
std_7th_delta_delta	0.0073	[MFCC]	[13]
mean_MFCC_2nd_coef	0.0068	[MFCC]	[14]
tqwt_entropy_log_dec_13	0.0067	[TQWT Features]	[15]
tqwt_entropy_shannon_dec_6	0.0066	[TQWT Features]	[16]
ddpJitter	0.0064	[Baseline Features]	[17]
std_10th_delta_delta	0.0063	[MFCC]	[18]
tqwt_entropy_shannon_dec_11	0.0062	[TQWT Features]	[19]
std_Log_energy	0.0057	[MFCC]	[20]

Table 6. Summary of cumulative weights and contribution ratios of core acoustic features for Parkinson’s disease
表 6. 帕金森语音核心特征：各大类累计权重与贡献度汇总

大类名称(Category)	特征数量	特征比例	累计权重	贡献占比
TQWT Features	153	64.3%	0.3802	56.68%
MFCC	33	13.9%	0.1759	26.22%
Wavelet Features	34	14.3%	0.0698	10.40%
Baseline Features	7	2.9%	0.0230	3.43%
Vocal Fold	7	2.9%	0.0144	2.14%

续表

Intensity Parameters	2	0.8%	0.0035	0.52%
Bandwidth Parameters	1	0.4%	0.0028	0.41%
Formant Frequencies	1	0.4%	0.0014	0.21%

2) 特征空间冗余度与判别独立性校验

如图 2 (相关性矩阵)所示, Top-10 核心特征间的相关性矩阵揭示了部分同源衍生特征间存在高度的局部共线性(Local Collinearity)。例如, std_6th_delta_delta 与 std_6th_delta 的皮尔逊相关系数高达 0.94, 提示在未来向边缘端轻量化部署时, 可通过流形降维进一步压榨冗余信息。但更为关键的是, 尽管特征内部存在局部强相关网络, 这些核心特征与目标诊断标签(class)之间均稳定保持了显著的独立临床判别相关性(相关系数绝对值分布在 0.26~0.36 之间)。这从统计学层面有力印证了: 即使在存在协同效应的特征簇中, 选定的核心声学指标依然能为疾病边界的划定提供不可替代的增量信息。

3) 全局特征权重的长尾稀疏分布验证

如图 2 (密度分布)所示, 全局特征重要性权重呈现出典型的极度右偏长尾分布(Right-skewed Long-tail Distribution)。绝大部分声学特征的权重极度收敛于 0 轴附近的分布峰部(Peak), 表明它们对分类决策实质上处于“休眠”或无效状态; 而极少数排名前列的核心提取指标则跨越了显著的统计学断层, 向高权重区间延伸, 构成了具有极高判别价值的长尾区域(Tail Region, 权重 >0.0100)。这一分布拓扑不仅在算法底层印证了高维临床声学特征的极度稀疏性(Sparsity), 更量化地确证了前文提取的 Top-20 核心声学参量正是主导模型决策的“绝对优势长尾”。求解过程将这一断崖式差异区域进行了量化标定, 通过计算各指标的统计学分布边界, 为剔除海量无关硬件噪声提供了坚实的依据, 从而确保了模型在真实医疗场景下的高灵敏度与高特异性。

4.4. 样本空间下的亚型分型可复现性检验

1) 亚型分类性能综合量化评估

将整合后的 221 名患者样本投射至统一的六维代理特征空间中, 后验逻辑斯蒂回归模型的全局评估结果如表 7 所示。在训练/验证集上的重复分层交叉验证中, 全局宏平均准确率达到 90.3%, 宏平均 F1 分数为 86.7%; 在完全未参与调参的独立测试集上, 准确率依然维持在 88.9%, 宏平均 F1 分数达 81.7%。这一严密的量化结果确凿地证明了, 本文依据“自适应双轨风险平面”提取的四类声学风险标签(混合型、不典型型、运动型风险、非运动型风险), 在底层统一降维空间中具备极高的特征可分性, 并在未知样本上展现出了卓越的泛化与复现能力。

Table 7. Classification performance evaluation metrics of the diagnostic model on cross-validation and independent test sets
表 7. 诊断模型在交叉验证与独立测试集上的分类性能评估指标

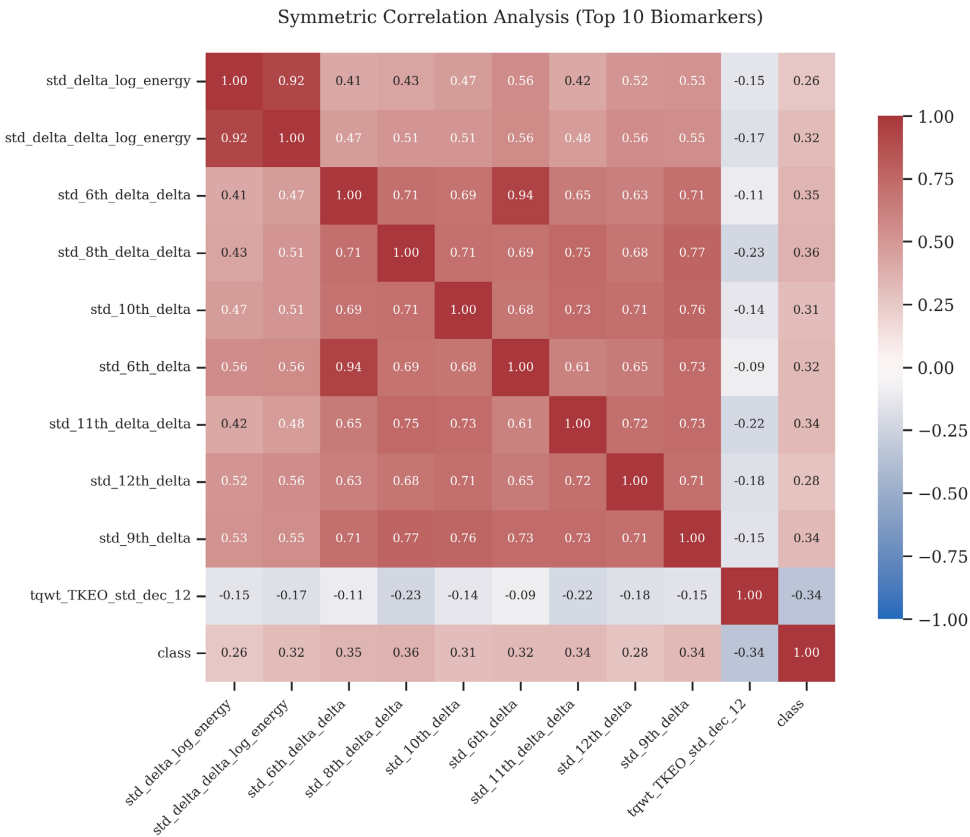
临床亚型类别	准确率(Acc)	精确率(Pre)	召回率(Rec)	F1 分数(F1)
Panel (a): 交叉验证集(Cross-validation Set, N = 176)				
运动型风险(Motor)	-	88.2%	92.3%	90.2%
非运动型风险(Non-Motor)	-	93.3%	97.2%	95.2%
混合型(Mixed)	-	87.5%	77.8%	82.4%
不典型型(Atypical)	-	88.2%	71.4%	78.9%
全局宏平均(Macro Avg)	90.3%	89.3%	84.7%	86.7%

续表

Panel (b): 独立测试集(Independent Test Set, N = 45)				
运动型风险(Motor)	-	94.1%	94.1%	94.1%
非运动型风险(Non-Motor)	-	89.5%	94.4%	91.9%
混合型(Mixed)	-	66.7%	40.0%	50.0%
不典型型(Atypical)	-	83.3%	100.0%	90.9%
全局宏平均(Macro Avg)	88.9%	83.4%	82.1%	81.7%

2) 决策边界流形与分类偏误结构深度剖析

为进一步白盒化剖析模型的决策边界与分类偏误结构(Misclassification Structure), 本文结合图 3 的混淆矩阵与表 7 的评估指标进行了深度比对。结果显示, “运动型”与“非运动型”风险展现出绝对的边界稳定性(独立测试集 F1 分数分别高达 94.1%与 91.9%, 召回率均>94%)。这在底层特征流形上印证了: 单向的局部构音肌群强直与中枢调控紊乱, 在六维代理空间中拥有极度清晰、近似正交的判别超平面。同时, “不典型型”100.0%的完美召回, 证明模型能敏锐剥离缺乏典型退行性发声表征的游离样本。相比之下, 极少数分类偏误集中于“混合型”样本(测试集召回率降至 40.0%), 错分样本呈现出向单一的“运动型”或“非运动型”降维归入的趋势(如图 3 所示)。这一偏误现象在医学病理与拓扑学上极具合理性: “混合型”本质为两类严重病理并发的叠加态(Superposition State), 其数据簇横跨两大主干流形的交汇边界, 导致决策边界具有内生的模糊性与交融性。总体而言, 主干类别极高的预测置信度, 充分证实了本文无监督声学分型逻辑的数理严谨性。



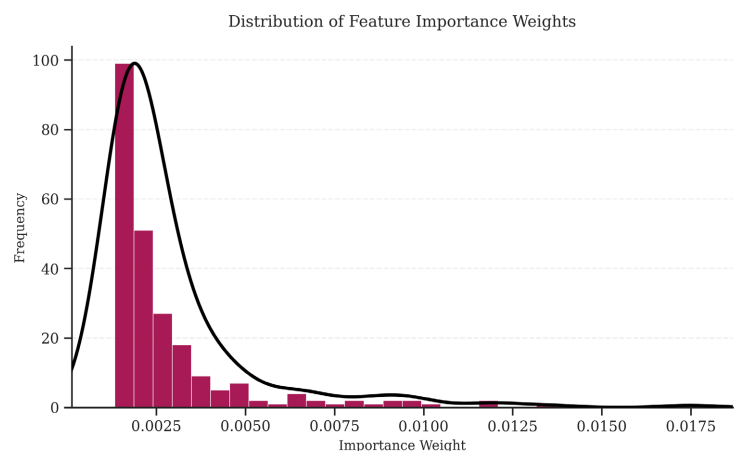
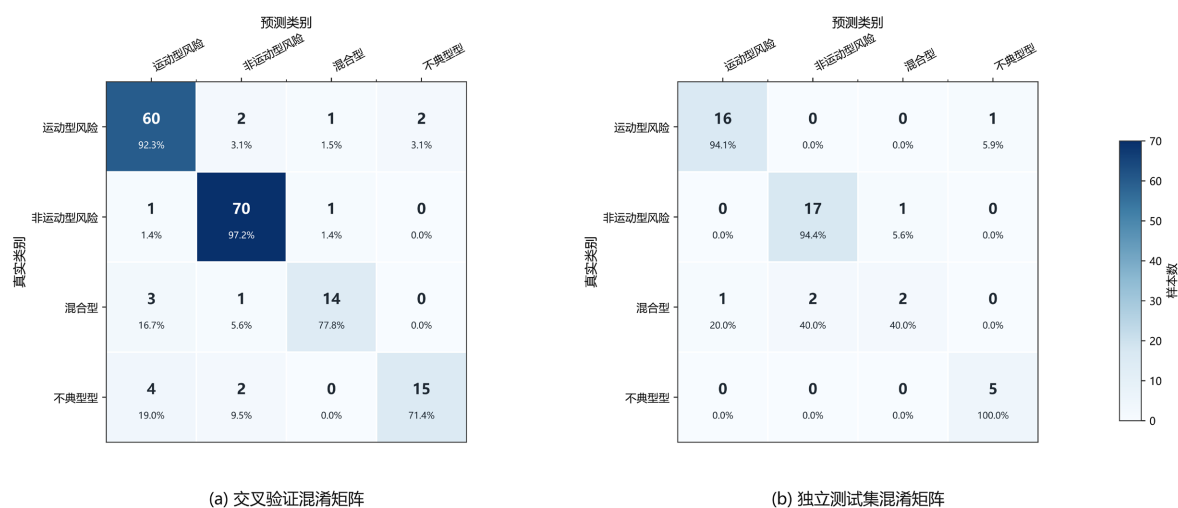


Figure 2. Correlation matrix of core acoustic features and global distribution of importance weights
图 2. 核心声学特征的相关性矩阵与重要性权重全局分布图



注：主对角线为预测正确样本；非对角线为错分样本。每格上方为样本数，下方为该真实类别内比例。

Figure 3. Confusion matrices of the four subtypes on the posterior validation set and test set
图 3. 四类亚型后验验证集及测试集混淆矩阵

5. 讨论

本研究针对帕金森病(PD)临床语音诊断中超高维特征冗余与缺乏显式亚型标签两大核心瓶颈，提出一套融合高维拓扑流形解构与自适应双轨风险映射的计算表型分析中枢。现就其方法学优势、局限性及未来转化方向探讨如下。

5.1. 方法学优势与理论意义

1) 非线性纯度增益驱动的物理降维机制：突破缺乏医学启发的线性降维局限，创新引入基于集成空间异质性散度衰减的特征筛选机制。通过量化基尼不纯度(Gini Impurity)下降量，揭示了高维声学特征的“重尾稀疏”分布规律。该机制有效消解了维数灾难，并将黑盒特征映射回具有明确医学解释的构音障碍指标(如 TQWT 与 MFCC)。

2) 宏微观诊断逻辑的拓扑连续性：通过发声解剖学机理先验，将降维后的异构数据集投射至六类声学代理空间，并顺滑过渡至双轨风险评分模型，使“底层特征提取”与“宏观医学表型”形成了严密、无缝的逻辑闭环。

3) 契合复杂临床并发症的无监督聚类拓扑：架构高度契合 PD 患者多系统病变、症状交织的真实流形。相较于传统硬聚类，本文构建的相对百分位(Percentile)映射平面在无标签环境下实现了精准的亚型界定，独立测试集宏平均 F1 分数达 81.7%，确证了其卓越的复现性与泛化能力。

5.2. 研究局限性与应用边界

医学先验依赖与长期临床验证的缺失：六维代理空间及双轨风险权重的构建，不可避免地包裹了较强的构音先验假设(如结构层权重及分型阈值的经验设定)。受限于公开基准库缺乏金标准，本文分型结果本质上仍属“特征流形层面的可复现性预警”，尚需大规模前瞻性临床队列的最终确证。

静态采样切片对疾病演进轨迹的局限：本研究依托特定诊断截面(Cross-sectional)的声学快照，尚未融入时间动力学(Temporal Dynamics)分析。PD 属动态退化过程，当前模型无法定量刻画患者亚型随病程演变的动态轨迹(如“不典型型”向其他亚型转化，或单一风险向“混合型”恶化)。

5.3. 未来工作与临床转化展望

联合多模态长程时序数据的流形重构：计划突破单次截面限制，引入多轮纵向随访语音序列，结合动态时间规整(DTW)等时序网络，构建发声退化的时间演化流形。进一步与步态、面部特征进行决策级融合，构建高维度立体数字生物标志物(Digital Biomarkers)体系。

普惠型边缘侧智能监测部署：本方法经非线性降维剔除冗余后算力开销极低，具备极高的边缘计算(Edge Computing)转化潜力。未来可封装为智能终端 App，高危人群仅需居家录制短促发声，即可实时输出风险雷达图，为神经退行性疾病的普惠型早筛与远程追踪提供低成本干预工具。

参考文献

- [1] Rabie, H. and Akhloufi, M.A. (2025) A Review of Machine Learning and Deep Learning for Parkinson's Disease Detection. *Discover Artificial Intelligence*, **5**, Article No. 24. <https://doi.org/10.1007/s44163-025-00241-9>
- [2] Rahman, S., Hasan, M., Sarkar, A.K. and Khan, F. (2023) Classification of Parkinson's Disease Using Speech Signal with Machine Learning and Deep Learning Approaches. *European Journal of Electrical Engineering and Computer Science*, **7**, 20-27. <https://doi.org/10.24018/ejece.2023.7.2.488>
- [3] Tsanas, A., Little, M.A., McSharry, P.E., Spielman, J. and Ramig, L.O. (2012) Novel Speech Signal Processing Algorithms for High-Accuracy Classification of Parkinson's Disease. *IEEE Transactions on Biomedical Engineering*, **59**, 1264-1271. <https://doi.org/10.1109/tbme.2012.2183367>
- [4] Tsanas, A., Little, M., McSharry, P. and Ramig, L. (2009) Accurate Telemonitoring of Parkinson's Disease Progression by Non-Invasive Speech Tests. *Nature Precedings*. <https://doi.org/10.1038/npre.2009.3920.1>
- [5] Karlsson, F., Schalling, E., Laakso, K., Johansson, K. and Hartelius, L. (2020) Assessment of Speech Impairment in Patients with Parkinson's Disease from Acoustic Quantifications of Oral Diadochokinetic Sequences. *The Journal of the Acoustical Society of America*, **147**, 839-851. <https://doi.org/10.1121/10.0000581>
- [6] Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., *et al.* (2013) Collinearity: A Review of Methods to Deal with It and a Simulation Study Evaluating Their Performance. *Ecography*, **36**, 27-46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>
- [7] Buda, M., Maki, A. and Mazurowski, M.A. (2018) A Systematic Study of the Class Imbalance Problem in Convolutional Neural Networks. *Neural Networks*, **106**, 249-259. <https://doi.org/10.1016/j.neunet.2018.07.011>
- [8] Zhang, M.-L. and Zhou, Z.-H. (2014) A Review on Multi-Label Learning Algorithms. *IEEE Transactions on Knowledge and Data Engineering*, **26**, 1819-1837. <https://doi.org/10.1109/tkde.2013.39>
- [9] Geurts, P., Ernst, D. and Wehenkel, L. (2006) Extremely Randomized Trees. *Machine Learning*, **63**, 3-42. <https://doi.org/10.1007/s10994-006-6226-1>

-
- [10] James, G., Witten, D., Hastie, T., Tibshirani, R. and Taylor, J. (2023) Statistical Learning. In: James, G., Witten, D., *et al.*, Eds., *Springer Texts in Statistics*, Springer International Publishing, 15-67.
https://doi.org/10.1007/978-3-031-38747-0_2
 - [11] Peng, X., Zhao, Y., Li, Z., Wang, X., Nan, F., Zhao, Z., *et al.* (2025) Multi-Scale and Multi-Level Feature Assessment Framework for Classification of Parkinson's Disease State from Short-Term Motor Tasks. *IEEE Transactions on Biomedical Engineering*, **72**, 1211-1224. <https://doi.org/10.1109/tbme.2024.3418688>
 - [12] Ryumina, E., Axyonov, A., Dolgushin, M., Ryumin, D. and Karpov, A. (2026) DEPART: Multi-Task Interpretable Depression and Parkinson's Disease Detection from In-The-Wild Video Data. *Big Data and Cognitive Computing*, **10**, Article 89. <https://doi.org/10.3390/bdcc10030089>
 - [13] Little, M., McSharry, P., Hunter, E., Spielman, J. and Ramig, L. (2008) Suitability of Dysphonia Measurements for Telemonitoring of Parkinson's Disease. *Nature Precedings*. <https://doi.org/10.1038/npre.2008.2298.1>
 - [14] Little, M.A., McSharry, P.E., Hunter, E.J., Spielman, J. and Ramig, L.O. (2009) Suitability of Dysphonia Measurements for Telemonitoring of Parkinson's Disease. *IEEE Transactions on Biomedical Engineering*, **56**, 1015-1022.
<https://doi.org/10.1109/tbme.2008.2005954>
 - [15] Tsanas, A., Little, M.A., McSharry, P.E. and Ramig, L.O. (2011) Nonlinear Speech Analysis Algorithms Mapped to a Standard Metric Achieve Clinically Useful Quantification of Average Parkinson's Disease Symptom Severity. *Journal of The Royal Society Interface*, **8**, 842-855. <https://doi.org/10.1098/rsif.2010.0456>
 - [16] Pandey, P.V.K., Sahu, S.S., Karan, B. and Mishra, S.K. (2024) Parkinson Disease Prediction Using CNN-LSTM Model from Voice Signal. *SN Computer Science*, **5**, Article No. 381. <https://doi.org/10.1007/s42979-024-02728-1>
 - [17] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L. and Polosukhin, I. (2017) Attention Is All You Need. arXiv: 1706.03762. <https://arxiv.org/abs/1706.03762>
 - [18] Gulati, A., Qin, J., Chiu, C., Parmar, N., Zhang, Y., Yu, J., *et al.* (2020) Conformer: Convolution-Augmented Transformer for Speech Recognition. *Interspeech 2020*, Shanghai, 25-29 October 2020, 5036-5040.
<https://doi.org/10.21437/interspeech.2020-3015>
 - [19] Schölkopf, B. and Smola, A.J. (2002) Learning with Kernels. The MIT Press.
<https://doi.org/10.7551/mitpress/4175.001.0001>
 - [20] Ruzs, J., Cmejla, R., Ruzickova, H. and Ruzicka, E. (2011) Quantitative Acoustic Measurements for Characterization of Speech and Voice Disorders in Early Untreated Parkinson's Disease. *The Journal of the Acoustical Society of America*, **129**, 350-367. <https://doi.org/10.1121/1.3514381>
 - [21] Postuma, R.B., Berg, D., Stern, M., Poewe, W., Olanow, C.W., Oertel, W., *et al.* (2015) MDS Clinical Diagnostic Criteria for Parkinson's Disease. *Movement Disorders*, **30**, 1591-1601. <https://doi.org/10.1002/mds.26424>
 - [22] Powers, D.M.W. (2020) Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation. arXiv: 2010.16061. <https://arxiv.org/abs/2010.16061>
 - [23] Zhou, Z.-H. (2018) A Brief Introduction to Weakly Supervised Learning. *National Science Review*, **5**, 44-53.
<https://doi.org/10.1093/nsr/nwx106>