研究生论文精粹

基于RTMPose和PatchTST的帕金森病和特发性震颤的视频鉴别诊断研究

彭宇盟[1,4,5]，于金泽[2]，潘隆盛[3]，曾梓敬[3,4]，袁　田[1,4]，时　颖[1,4]，张政波[1]
[1]解放军总医院医学创新研究部，北京　100853；[2]北京航空航天大学计算机学院，北京　100083；[3]解放军总医院第一医学中心神经外科医学部，北京　100853；[4]解放军医学院，北京　100853；[5]联勤保障部队第923医院神经内科，广西南宁　530021

摘要：背景　帕金森病(Parkinson's disease，PD)与特发性震颤(essential tremor，ET)临床表现相似，当前诊断依赖神经科医师主观量表评估，耗时长且一致性有限。**目的**　开发一种基于视频分析的智能分类模型，结合深度学习实现PD与ET的高效自动鉴别，为无创诊断提供新思路。**方法**　纳入2021年11月—2024年1月解放军总医院门诊14例PD患者与63例ET患者，采集其执行3种标准化上肢运动任务(手指指鼻、翻手掌、握拳张开)的1 136段视频。基于MMPose框架的RTMPose模型提取手腕及手指关键点坐标序列，计算位移、速度及加速度等运动学特征，构建包含时空轨迹与统计学特征数据集。以Transformer架构建立PatchTST模型(输入特征序列按时间窗口分块处理，融合全局注意力机制)，并与逻辑回归、XGBoost、随机森林、支持向量机、Informer及长短期记忆网络进行对比。**结果**　PatchTST模型在融合关键点坐标与运动学特征时的平均模型性能最优，其在手指指鼻任务中的准确度最佳(AUC=0.957)，3种运动任务的平均AUC达到了0.897。在全部21种模型组合中，仅纳入运动学特征的LSTM模型性能最差，3种运动任务的平均AUC仅为0.691。**结论**　基于视频的PD与ET智能鉴别诊断方法依托人体姿态估计与深度学习技术，能够高精度、高效率实现无接触的远程诊断，为运动障碍疾病的早期诊断和管理提供参考。

关键词：帕金森病；特发性震颤；视频诊断；人体姿态估计；深度学习；Transformer模型；远程医疗
中图分类号：R742. 5　　**文献标志码：**A　　**文章编号：**2095-5227(2025)07-0638-08
DOI：10.12435/j.issn.2095-5227.25010501
引用本文：彭宇盟，于金泽，潘隆盛，等.基于RTMPose和PatchTST的帕金森病和特发性震颤的视频鉴别诊断研究[J].解放军医学院学报，2025，46（7）：638-645.

Video-based differential diagnosis of Parkinson's disease and essential tremor using RTMPose and PatchTST

PENG Yumeng[1,4,5], YU Jinze[2], PAN Longsheng[3], ZENG Zijing[3,4], YUAN Tian[1,4], SHI Ying[1,4], ZHANG Zhengbo[1]
[1]Medical Innovation Research Department of PLA General Hospital, Beijing 100853, China; [2]School of Computer Science and Engineering, Beihang University, Beijing 100083, China; [3]Department of Neurosurgery, the First Medical Center of PLA General Hospital, Beijing 100853, China; [4]Chinese PLA Medical School, Beijing 100853, China; [5]Department of Neurology, 923th Hospital of the Joint Logistics Support Force of PLA, Nanning 530021, Guangxi Zhuang Autonomous Region, China
Corresponding authors: ZHANG Zhengbo. Email: zhengbozhang@126.com

Abstract: Background　Parkinson's disease (PD) and essential tremor (ET) share overlapping clinical manifestations. Current diagnostic approaches rely on subjective rating scales by neurologists, which are time-consuming and limited inter-rater consistency. **Objective**　To develop an intelligent classification model based on video analysis, integrating deep learning to achieve efficient and automated differentiation between PD and ET, thereby offering a novel approach for non-invasive diagnosis. **Methods**　A total of 14 PD and 63 ET patients from the Outpatient Department of PLA General Hospital (Nov.2021 to Jan.2024) were enrolled. A dataset comprising 1 136 video clips was collected during the performance of three standardized upper limb motor tasks: finger-to-nose, hand pronation-supination, and fist opening-closing. Using the RTMPose model within the MMPose framework, keypoint coordinates of the wrist and fingers were extracted. Kinematic features such as displacement, velocity, and acceleration were computed to construct a dataset of spatiotemporal trajectories and statistical descriptors. A Transformer-based PatchTST model was

收稿日期：2025-01-05
基金项目：北京市自然科学基金资助项目(7252299)
第一作者：彭宇盟，在读硕士，医师。Email: wspym@foxmail.com
通信作者：张政波，博士，高级工程师。Email: zhengbozhang@126.com

developed, in which temporal sequences were segmented into patches and processed via global attention mechanisms. Model performance was compared against logistic regression, XGBoost, random forest, support vector machine, Informer, and long short-term memory (LSTM) networks. **Results** The PatchTST model achieved the best average performance when combining keypoint coordinates with kinematic features. The highest accuracy was observed in the finger-to-nose task (AUC=0.957), with an overall average AUC of 0.897 among all three tasks. Among the 21 model-feature combinations, the LSTM model using only kinematic features performed the worst, with an average AUC of 0.691. **Conclusion** The video-based intelligent differential diagnosis method for PD and ET, leveraging human pose estimation and deep learning technologies, enables high-precision and high-efficiency remote diagnosis without physical contact, providing reference for early diagnosis and management of movement disorders.

Keywords: Parkinson's disease; essential tremor; video-based diagnosis; human pose estimation; deep learning; transformer model; telemedicine

Cited as: Peng YM, Yu JZ, Pan LSH, et al. Video-Based differential diagnosis of Parkinson's disease and essential tremor using RTMPose and PatchTST[J]. Acad J Chin PLA Med Sch, 2025, 46(7): 638-645.

帕金森病(Parkinson’s disease，PD)是最常见的神经退行性疾病，我国65岁以上老年人群中帕金森病的患病率约为1.7%，80岁以上超过4%[1]。特发性震颤(essential tremor，ET)是发病率最高的运动障碍疾病，其发病率随年龄增长而增加，在65岁以上老年人群中达到5.79%[2]。PD主要以静止性震颤为主，ET主要以运动性震颤为主，两者症状存在一定的重叠性，如震颤和步态异常，因此在鉴别诊断上较为困难。目前对这两种疾病的诊断主要依赖神经科医师通过临床评定量表进行面对面评估，这种方法主观性强、耗时长且复杂[3]。一项研究报告显示，约1/3的PD患者被误诊为ET，导致这些患者未能及时接受适当的治疗[4]。

为了克服传统诊断方法的不足，基于传感器的智能诊断方法应运而生[5]。此类方法利用人工智能对患者的运动数据进行分析，从而提取肢体的运动学特征，实现对运动障碍疾病的客观、准确诊断和评估，目前已有不少基于可穿戴设备对PD和ET进行诊断和鉴别的研究[6-7]。基于视频的鉴别诊断技术无需任何特殊设备，仅通过电脑或智能手机摄像头即可实现远程、无接触、无创诊断[5]。由于基于视频的方法相比可穿戴传感器更难提取运动特征，同时模型性能对结果的影响较大，目前相关研究较少且准确度低[8-9]。本研究旨在开发一种基于上肢运动视频分析的智能诊断方法，结合人体姿态估计与深度学习技术，实现PD与ET的高效鉴别，克服传统方法的局限性，提升诊断准确性与临床应用可行性。

1 对象与方法

1.1 研究对象

收集2021年11月19日—2024年1月16日在解放军总医院门诊就诊的患者。纳入标准：(1)PD患者符合UK脑库PD临床诊断标准，包括运动减少、至少1个附加症状(肌肉强直、静止性震颤或直立不稳)，以及3个或以上确诊标准(如单侧起病、存在静止性震颤、疾病逐渐进展等)；(2)ET患者符合美国运动障碍学会及世界震颤研究组织提出的ET诊断标准，包括核心诊断标准(如双手及前臂动作性震颤、不伴有其他神经系统体征等)和次要诊断标准(如病程超过3年、有家族史等)；(3)有能力并愿意参加所有研究访视；(4)无认知障碍，能够在治疗期间与医师沟通感受；(5)无重度焦虑、抑郁；无严重心血管疾病、脑血管疾病。本研究已获得解放军总医院伦理委员会批准(S2018-021-00/01)，所有受试者均签署知情同意书。

1.2 视频任务选择与设计

参考国际帕金森与运动障碍震颤工作组撰写的关于震颤分类的共识声明[10]，选用临床评估PD最常用的统一帕金森病评定量表(Unified Parkinson's Disease Rating Scale，UPDRS)[11]和评估ET最常用的Fahn-Tolosa-Marin震颤评定量表(Fahn-Tolosa-Marin Tremor Rating Scale，FTM-TRS)[12]作为任务设计依据。通过总结UPDRS量表的第3部分(运动功能评估)和FTM-TRS量表中与震颤评估相关的运动任务，我们最终选择了两个量表中均包含的3种上肢运动任务进行视频采集，即双手动作性震颤(手指指鼻)、双手前臂回旋运动(翻手掌)、双手手掌运动(握拳张开)，见图1。受试者的视频数据在神经科医师的指导下采集。受试者以舒适的姿势坐在椅子上，并根据FTM-CRST量表的标准，在神经科医师的引导下执行上述3种运动任务。采集过程通过小米平板的后置摄像头全程记录。最终，从上述3种任务中共收集到1 136个视频(视频分辨率为1 920×1 080，帧率为30 fps，格式为mp4)。

1.3 软件环境及研究框架

采用AMD Ryzen 9 5950X CPU和NVIDIA



图 1　本研究纳入的3种上肢运动任务

Fig. 1　Three types of upper limb motor tasks included in this study

GeForce RTX 3090 GPU进行数据处理、模型训练与测试。使用PyTorch 2.5.1和CUDA 12.4进行深度学习模型的训练与推理，并基于MMPose 1.2.0框架对人体姿态估计模型提取PD与ET患者在运动任务视频中手腕和手指的关键点坐标，生成手部运动的关键点序列及统计学特征。采用基于Transformer架构的时间序列预测与分类模型对这些关键点序列进行分类分析。模型架构见图2。

1.4　手部关键点坐标序列和运动特征提取

我们调研了多种人体姿态估计方法(Paddle Detection[13]、AlphaPose[14]、MMPose[15])与不同的检测器(YOLOv3[16]、Faster-RCNN[17]和RTMDet[18])结合的姿态估计性能，经过比较，我们发现基于MMPose框架的RTMPose(Real-Time Multi-Person Pose Estimation based on MMPose)模型[19]在平衡模型性能和复杂性方面表现最佳，并显示出较强的检测鲁棒性。

最终我们选取效果最佳的RTMPose模型应用于患者的运动任务视频进行人体姿态估计，并提取视频中每1帧图像的人体关键点坐标。对所选择的3种评估运动任务——双手动作性震颤(手指指鼻)、双手手掌运动(握拳张开)和双手前臂回旋运

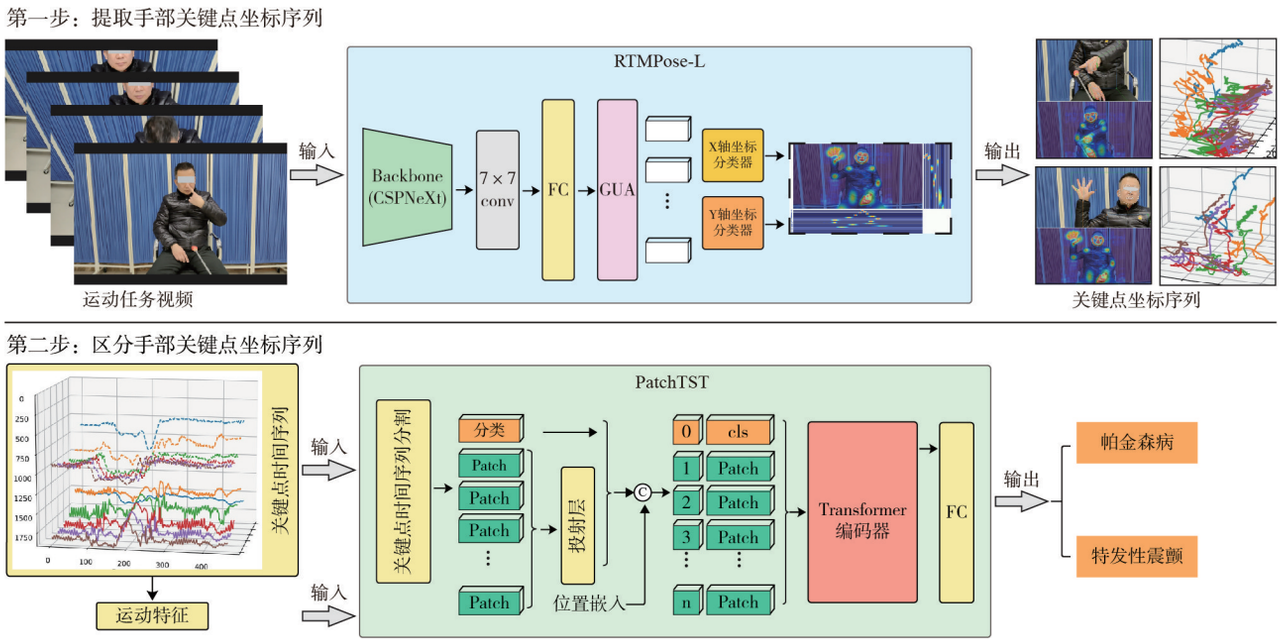


图 2　基于视频的帕金森病和特发性震颤的智能辅助诊断研究模型架构

Fig. 2　Model framework for video-based intelligent auxiliary diagnosis of Parkinson's disease and essential tremor

动(翻手掌)，分别以左手和右手作为评估对象，对视频中的每1帧图像进行全身姿态估计并提取目标手的手腕和手指关键点坐标，形成共计8组手部关键点坐标序列，每组坐标序列的维度为6(分别代表手腕和5个手指的坐标)，输出结果见图3。接下来，对于其中的每组坐标序列，我们将RTMPose-L模型输出的绝对坐标转换为相对坐标，即每个手部关键点的坐标减去其在第1帧中的坐标，得到表示位移的相对坐标。使用得到的相对坐标序列，计算速度、加速度、振幅、频率和熵5种坐标序列上的统计特征，并将坐标序列与从坐标序列上得到的统计特征保存，作为下一步特征提取和分析的输入。见图3。

1.5　手部关键点坐标序列分类

我们采用PatchTST (Patch Time Series Transformer)模型[20]对手部关键点坐标序列进行分类。用于分类的PatchTST模型首先将输入序列划分为多个长度为P的补丁块，通过Transformer的编码器对每个输入补丁进行特征提取并输出补丁预测，最后将输出结果展平为一维向量，并通过全连接层进行PD与ET的分类。

我们采用五折交叉验证将所有手部关键点坐标序列按患者编号分组，每组中的坐标序列来自同一患者，然后将这些分组划分为5个互斥的子集，每次使用其中4个子集作为训练集，剩余1个子集作为验证集，重复5次，每次使用不同的验证集进行训练与测试。

1.6　模型性能对比设置

为了进一步验证本研究方法的有效性，我们选择了基于Transformer架构的时间序列分类模型PatchTST，Kovalenko等[8]、Hayashida等[9]使用的逻辑回归、XGBoost、随机森林和支持向量机模型，以及常用的长短期记忆网络模型(long short-term memory，LSTM)和Informer模型，进行对比分析，并计算准确率(accuracy，ACC)、曲线下面积(area under curve，AUC)、F1值(F1 Score)评估模型的分类性能。

2　结果

2.1　受试者信息

截至2024年1月16日，共纳入PD患者14例，ET患者63例。其中PD组包含男性10例，女性4例，平均年龄(68±7)岁(范围：53～78岁)，中位年龄68(*IQR*：67～74)岁，视频数据量(mp4) 553。ET组包含男性46例，女性17例，平均年龄(64±9)岁(范围：30～80岁)，中位年龄66(*IQR*：61～70)岁，视频数据量(mp4) 583。

2.2　模型鉴别性能

本研究通过五折交叉验证评估了不同特征组合与模型在3种上肢运动任务中的分类性能。结果

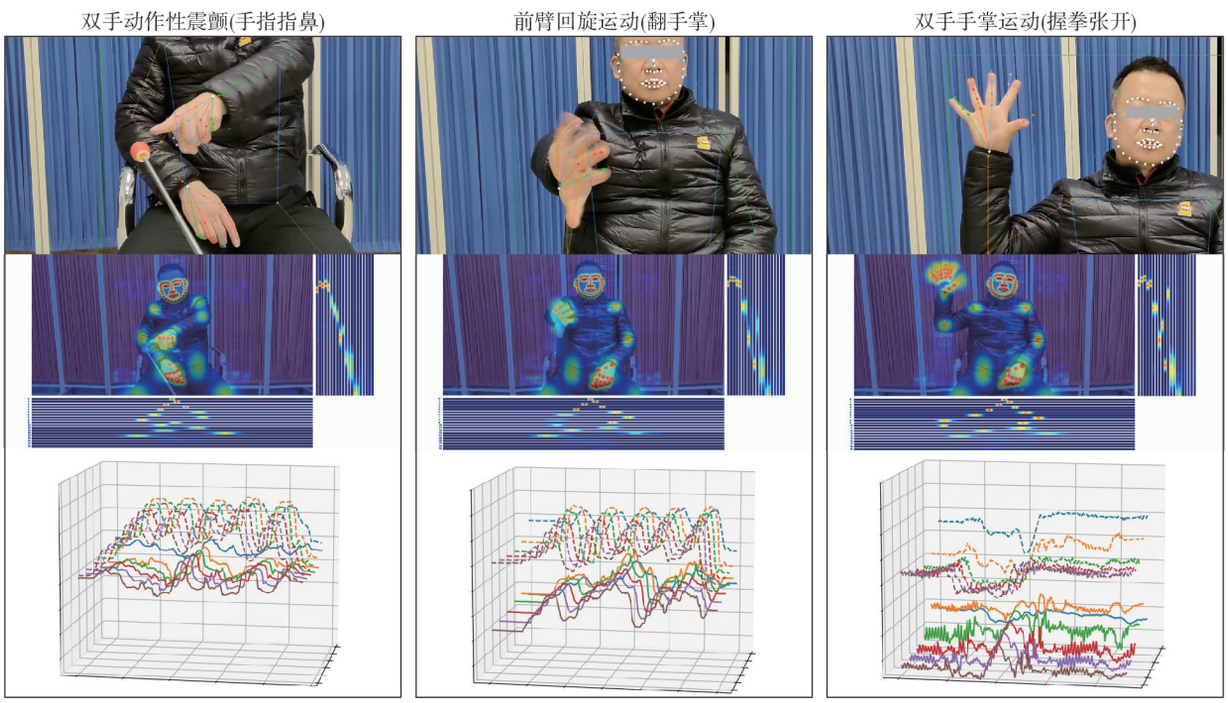


图3　使用RTMPose-L模型提取的人体关键点

Fig. 3　Human key points extracted using the RTMPose-L model

显示，基于Transformer架构的PatchTST模型在融合手部关键点坐标序列与统计学运动特征时表现最优。尤其在双手动作性震颤(手指指鼻)任务中显著优于其他模型。见表1～表4。

在双手动作性震颤(手指指鼻)任务中，PatchTST模型在融合手部关键点坐标序列与统计学运动特征组合下的性能最佳，较传统机器学习模型提升超过12%。对比基线实验发现，传统机器学习模型如逻辑回归、支持向量机(Support Vector Machine，SVM)在处理高维时序数据时性能受限，而LSTM因梯度问题表现最差。对3种运动任务的模型性能结果分析发现，双手动作性震颤(手指指鼻)任务因动作简单且震颤特征显著，模型整体性能最佳。前臂回旋运动(翻手掌)任务的患者姿态变异性大，导致模型性能下降。

通过分析发现，基于x、y坐标的手部关键点序列特征直接反映手部运动轨迹的时空动态，但缺乏对运动特性的量化描述，统计学运动特征(速度、加速度、振幅、频率和熵)的动态指标信息密度较低。两者均低于其在关键点坐标序列与统计学运动特征组合下的准确度，说明单一统计特征难以表征复杂运动模式。而手部关键点坐标序列与统计学运动特征结合时，融合了时空轨迹与动态特性，模型的分类效果最佳。

2.3　PatchTST模型性能的可视化

我们绘制了PatchTST模型在3种运动任务中融合手部关键点坐标序列与统计学运动特征的混淆矩阵和ROC曲线，见图4。结果显示，模型在双手动作性震颤(手指指鼻)任务中表现出色，混淆矩阵显示分类准确率较高，误分类样本较少。ROC曲线的AUC值接近1，在不同阈值下具有良好的区分能力。各折线的AUC值均较高，表明模型在不同数据子集上的表现稳定。双手前臂回旋运动(翻手掌)任务中混淆矩阵显示50个假阳性以及40个假阴性。与动作性震颤任务相比，这一任务的分类表现略逊一筹，假阳性数量较高。ROC曲线不同折的AUC值范围0.764～0.934，各折线的AUC值存在一定的波动，模型在不同数据子集上的表现不够稳定，但平均值仍达到了0.816。双手手掌运动(握拳张开)任务中的模型表现良好，预测和实际值的分布合理，具有较高的分类准确率。

3　讨论

本研究提出了一种基于视频分析的智能诊断

表1　模型在双手动作性震颤(手指指鼻)任务的分类性能评估结果($\bar{x}\pm s$)

Tab. 1　Classification performance evaluation results of each model on bilateral action tremor (finger-to-nose) ($\bar{x}\pm s$)

特征组合	模型	ACC	AUC	F1值	模型性能/%	性能排序
坐标序列	PatchTST	0.897±0.048	0.947±0.039	0.894±0.052	-1.88	2
	Informer	0.864±0.061	0.930±0.048	0.832±0.062	-5.92	5
	LSTM	0.777±0.077	0.837±0.067	0.721±0.118	-16.36	16
	Logistic Regression	0.779±0.041	0.850±0.037	0.764±0.059	-14.27	15
	Random Forest	0.779±0.027	0.852±0.038	0.765±0.043	-14.17	14
	XGBoost	0.801±0.027	0.874±0.043	0.794±0.031	-11.55	11
	SVM	0.796±0.025	0.861±0.024	0.782±0.034	-12.62	13
运动特征	PatchTST	0.826±0.042	0.865±0.036	0.825±0.036	-9.82	10
	Informer	0.860±0.051	0.930±0.039	0.859±0.053	-5.09	4
	LSTM	0.638±0.083	0.623±0.138	0.476±0.246	-37.79	21
	Logistic Regression	0.738±0.030	0.808±0.046	0.719±0.040	-18.87	18
	Random Forest	0.746±0.044	0.840±0.044	0.727±0.043	-17.17	17
	XGBoost	0.721±0.020	0.824±0.035	0.705±0.010	-19.44	19
	SVM	0.699±0.034	0.801±0.023	0.671±0.043	-22.27	20
坐标序列+运动特征	PatchTST	0.917±0.041	0.957±0.028	0.916±0.041	0.00	1
	Informer	0.896±0.052	0.933±0.080	0.880±0.040	-2.91	3
	LSTM	0.812±0.096	0.861±0.090	0.780±0.122	-12.11	12
	Logistic Regression	0.850±0.027	0.916±0.022	0.833±0.041	-6.88	7
	Random Forest	0.814±0.027	0.922±0.023	0.786±0.059	-9.69	9
	XGBoost	0.851±0.027	0.933±0.016	0.836±0.053	-6.15	6
	SVM	0.825±0.051	0.911±0.030	0.795±0.070	-9.35	8

表 2 模型在前臂回旋运动(翻手掌)任务上的分类性能评估结果($\bar{x}\pm s$)

Tab. 2 Classification performance evaluation results of each model on forearm rotation (pronation-supination of the palms) ($\bar{x}\pm s$)

特征组合	模型	ACC	AUC	F1值	模型性能/%	性能排序
坐标序列	PatchTST	0.798±0.039	0.837±0.054	0.767±0.032	2.73	13
	Informer	0.800±0.051	0.896±0.038	0.797±0.053	6.56	4
	LSTM	0.695±0.083	0.721±0.092	0.657±0.141	-11.34	19
	Logistic Regression	0.811±0.031	0.890±0.031	0.803±0.023	7.05	3
	Random Forest	0.786±0.026	0.878±0.032	0.773±0.042	4.15	11
	XGBoost	0.801±0.029	0.880±0.035	0.789±0.025	5.59	7
	SVM	0.770±0.035	0.864±0.026	0.747±0.066	1.74	14
运动特征	PatchTST	0.703±0.054	0.734±0.083	0.705±0.045	-8.34	17
	Informer	0.800±0.032	0.863±0.055	0.799±0.031	5.30	8
	LSTM	0.636±0.059	0.656±0.062	0.562±0.131	-20.75	20
	Logistic Regression	0.579±0.088	0.607±0.077	0.575±0.083	-24.65	21
	Random Forest	0.783±0.060	0.860±0.035	0.777±0.044	3.46	12
	XGBoost	0.767±0.028	0.847±0.044	0.761±0.027	1.53	15
	SVM	0.685±0.057	0.736±0.058	0.668±0.069	-10.67	18
坐标序列+运动特征	PatchTST	0.767±0.036	0.816±0.041	0.755±0.035	0.00	16
	Informer	0.832±0.030	0.896±0.036	0.814±0.022	8.70	1
	LSTM	0.791±0.037	0.855±0.030	0.793±0.032	4.31	10
	Logistic Regression	0.818±0.057	0.860±0.032	0.804±0.060	6.18	5
	Random Forest	0.803±0.052	0.883±0.030	0.787±0.061	5.71	6
	XGBoost	0.814±0.022	0.883±0.014	0.810±0.059	7.21	2
	SVM	0.804±0.021	0.875±0.021	0.783±0.061	5.25	9

表 3 模型在双手手掌运动(握拳张开)任务上的分类性能评估结果($\bar{x}\pm s$)

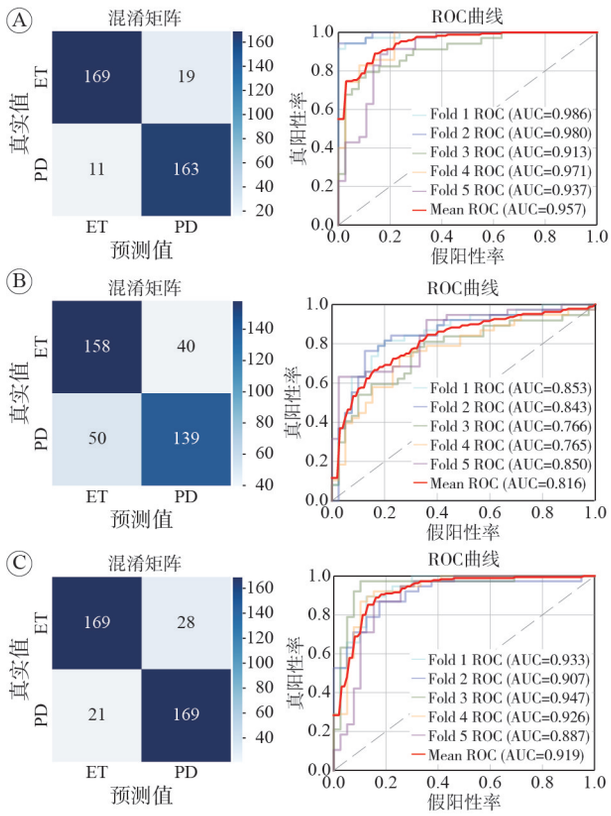
Tab. 3 Classification performance evaluation results of each model on Bilateral palm movement (fist-clenching and opening) ($\bar{x}\pm s$)

特征组合	模型	ACC	AUC	F1值	模型性能/%	性能排序
坐标序列	PatchTST	0.868±0.030	0.912±0.032	0.845±0.033	-1.51	2
	Informer	0.780±0.069	0.858±0.042	0.773±0.076	-9.58	11
	LSTM	0.695±0.078	0.739±0.052	0.648±0.221	-21.92	21
	Logistic Regression	0.754±0.073	0.836±0.052	0.766±0.072	-11.64	13
	Random Forest	0.726±0.018	0.816±0.032	0.716±0.025	-15.34	17
	XGBoost	0.760±0.035	0.825±0.024	0.752±0.037	-12.34	14
	SVM	0.711±0.032	0.801±0.040	0.710±0.030	-16.69	19
运动特征	PatchTST	0.816±0.030	0.860±0.034	0.821±0.033	-6.30	3
	Informer	0.784±0.048	0.874±0.050	0.783±0.049	-8.47	8
	LSTM	0.755±0.049	0.794±0.071	0.737±0.084	-14.23	16
	Logistic Regression	0.698±0.082	0.753±0.067	0.684±0.095	-19.92	20
	Random Forest	0.791±0.020	0.862±0.016	0.790±0.018	-8.37	7
	XGBoost	0.773±0.019	0.848±0.034	0.771±0.017	-10.29	12
	SVM	0.726±0.03	0.798±0.032	0.731±0.029	-15.42	18
坐标序列+运动特征	PatchTST	0.873±0.022	0.919±0.021	0.873±0.022	0.00	1
	Informer	0.764±0.080	0.910±0.031	0.790±0.049	-7.66	5
	LSTM	0.765±0.081	0.810±0.096	0.741±0.106	-13.12	15
	Logistic Regression	0.774±0.053	0.855±0.032	0.787±0.065	-9.39	10
	Random Forest	0.783±0.046	0.884±0.025	0.780±0.063	-8.26	6
	XGBoost	0.803±0.021	0.883±0.024	0.789±0.063	-7.19	4
	SVM	0.793±0.021	0.844±0.021	0.790±0.062	-8.94	9

表 4　模型总体分类性能评估结果

Tab. 4　Average classification performance evaluation results of the models

特征组合	模型	平均模型性能/%	平均模型性能排序
坐标序列	PatchTST	-0.22	2
	Informer	-2.98	6
	LSTM	-16.54	19
	Logistic Regression	-6.29	11
	Random Forest	-8.45	15
	XGBoost	-6.10	10
	SVM	-9.19	16
运动特征	PatchTST	-8.15	14
	Informer	-2.75	5
	LSTM	-24.26	21
	Logistic Regression	-21.15	20
	Random Forest	-7.36	13
	XGBoost	-9.40	17
	SVM	-16.12	18
坐标序列+运动特征	PatchTST	0.00	1
	Informer	-0.62	3
	LSTM	-6.97	12
	Logistic Regression	-3.36	7
	Random Forest	-4.08	8
	XGBoost	-2.04	4
	SVM	-4.35	9



A：双手动作震颤；B：双手翻手掌；C：双手指掌屈伸运动。

图 4　PatchTST 模型对 3 种上肢运动任务中 PD 和 ET 进行分类的混淆矩阵和 ROC 曲线

Fig. 4　Confusion matrix and ROC curves for classifying PD and ET based on three upper limb motor tasks

框架，通过融合人体姿态估计与深度学习，实现了 PD 与 ET 的客观精准鉴别诊断，在双手动作性震颤任务中 AUC 达到了 0.957。相较于传统依赖主观量表或可穿戴设备的鉴别方法，该方案仅需普通摄像头即可捕捉患者上肢运动特征，在提升诊断客观性的同时降低了技术门槛。

目前 PD 与 ET 的鉴别主要依赖神经科医师对临床症状的主观评估(如 UPDRS、FTM-TRS 等)，该过程依赖较强的经验与培训背景，量表间重测信度和评分者间一致性有限。此外，某些症状如姿势性震颤或起始动作迟缓在两类疾病中均有表现，进一步增加了临床判断的困难。尽管近年来可穿戴设备(如惯性测量单元 IMU)在识别运动障碍方面展现出一定潜力，但其设备成本、使用依从性以及对患者日常行为的干扰仍是主要限制。英国国家健康与护理卓越研究所 NICE 2023 年指南已明确推荐可穿戴技术用于 PD 监测[21]，而本研究通过非接触式视频分析进一步拓展了这一方向的应用场景，尤其适用于医疗资源匮乏地区或行动不便的老年患者。

截至 2025 年 2 月 10 日，我们在 Web of Science、EI、PubMed 等多个数据库检索了以 PD 与 ET 视频鉴别诊断为主题的相关研究，最终仅筛选出 2 篇基于视频对 PD 和 ET 进行鉴别的研究。现有研究在数据采集与模型泛化性上面临显著挑战。如 Hayashida 等[9]要求患者在绿幕前保持 30 cm 的距离拍摄，其复杂操作流程严重限制临床推广，且单一任务设计导致分类准确率仅为 0.56；Kovalenko 等[8]采用 OpenPose 框架提取关键点，尽管在低端设备上兼容性较好，但其算法精度与模型灵活性(如检测器不可调整)已落后于最新技术。这些研究多数采用静态图像或有限动作片段，忽略了震颤节律和动作序列的时间依赖性，对震颤频率、幅度及波动性的建模能力有限，因此无法有效表征 PD 与 ET 在运动控制机制上的核心差异。

本研究通过 RTMPose-L 模型实现高精度关键点检测，并创新性地将手部关键点坐标转换为相对位移、速度、加速度及运动熵等多维度时序特征。结合 PatchTST 模型的自注意力机制，系统可有效捕捉震颤节律异常和运动协调性障碍的长程依赖关系，最终在指鼻试验任务中达到 0.917 的准

确率与0.957的AUC，较Kovalenko等[8]报告的0.770的准确率提升19%。本研究设计了3种标准化上肢运动任务(指鼻试验、握拳张开、前臂旋前旋后)，通过多任务特征融合增强模型对异质性症状的识别能力。研究结果表明，即便在计算资源受限的场景下，XGBoost模型仍能在握拳张开任务中维持0.883的准确率。此外，基于关键点坐标的分析策略规避了背景、光照或服装等混杂因素的干扰，无需特殊拍摄设备即可保证结果稳定性，较传统视频分析方法更具普适性，提高了基于该方法的设备在基层医疗机构落地的可行性。

此外，本研究也有一定的局限性。当前研究受限于PD患者样本量，未对震颤主导型与非震颤亚型进行分层分析。未来工作将联合神经科医师构建更大规模的多中心数据集，并探索可解释性AI技术(如注意力权重可视化与震颤频谱分析相结合的策略)，以增强诊断透明度和亚型鉴别能力。

综上，本研究已验证基于普通视频的运动分析技术在PD/ET鉴别中的可行性，弥补了现有主观量表及可穿戴技术在临床应用中的缺陷，为运动障碍疾病的远程化、标准化评估提供了新范式。

作者贡献 彭宇盟：课题设计，数据采集，文章撰写；于金泽：模型构建；曾梓敬、袁田、时颖：数据采集；潘隆盛：临床指导；张政波：课题及模型指导。

利益冲突 所有作者声明不存在利益冲突。

数据共享声明 本研究相关数据可根据合理要求，通过联系通信作者获取。

参考文献

1 Zhu JQ，Cui YS，Zhang JJ，et al. Temporal trends in the prevalence of Parkinson’s disease from 1980 to 2023：a systematic review and meta-analysis［J］. Lancet Healthy Longev，2024，5（7）：e464-e479.

2 Louis ED，Ottman R，Hauser WA. How common is the most common adult movement disorder? estimates of the prevalence of essential tremor throughout the world［J］. Mov Disord，1998，13（1）：5-10.

3 Jankovic J，Tan EK. Parkinson’s disease：etiopathogenesis and treatment［J］. J Neurol Neurosurg Psychiatry，2020，91（8）：795-808.

4 Moon S，Song HJ，Sharma VD，et al. Classification of Parkinson’s disease and essential tremor based on balance and gait characteristics from wearable motion sensors via machine learning techniques：a data-driven approach［J］. J Neuroeng Rehabil，2020，17（1）：125.

5 Peng YM，Ma CB，Li MW，et al. Intelligent devices for assessing essential tremor：a comprehensive review［J］. J Neurol，2024，271（8）：4733-4750.

6 Ma CB，Li DY，Pan LS，et al. Quantitative assessment of essential tremor based on machine learning methods using wearable device［J］. Biomed Signal Process Control，2022，71：103244.

7 Hathaliya JJ，Modi H，Gupta R，et al. Parkinson and essential tremor classification to identify the patient’s risk based on tremor severity［J］. Comput Electr Eng，2022，101：107946.

8 Kovalenko E，Talitckii A，Anikina A，et al. Distinguishing between Parkinson’s disease and essential tremor through video analytics using machine learning：a pilot study［J］. IEEE Sens J，2021，21（10）：11916-11925.

9 Hayashida T，Sugiyama T，Sakai K，et al. Validation and discussion of severity evaluation and disease classification using tremor video［J］. Electronics，2023，12（7）：1674.

10 Bhatia KP，Bain P，Bajaj N，et al. Consensus Statement on the classification of tremors. from the task force on tremor of the International Parkinson and Movement Disorder Society［J］. Mov Disord，2018，33（1）：75-87.

11 Goetz CG，Tilley BC，Shaftman SR，et al. Movement disorder society-sponsored revision of the unified Parkinson’s disease rating scale （MDS-UPDRS）：scale presentation and clinimetric testing results［J］. Mov Disord，2008，23（15）：2129-2170.

12 Fahn S，Tolosa E，Marin C. Clinical Rating Scale for Tremor［M］//Jankovik J，Tolosa E. Parkinson's Disease and Movement Disorders. Baltimore-Münich：Urban & Schwarzenberg，1988：225-234.

13 PaddleDetection：Object Detection toolkit based on PaddlePaddle［EB/OL］. https://github.com/PaddlePaddle/PaddleDetection.

14 Fang HS，Li JF，Tang HY，et al. AlphaPose：whole-body regional multi-person pose estimation and tracking in real-time［J］. IEEE Trans Pattern Anal Mach Intell，2023，45（6）：7157-7173.

15 MMPose Contributors. Openmmlab pose estimation toolbox and benchmark［EB/OL］. https://github.com/openmmlab/mmpose.

16 Redmon J，Farhadi A. YOLOv3：an incremental improvement［J/OL］. https://doi.org/10.48550/arXiv.1804.02767.

17 Ren SQ，He KM，Girshick R，et al. Faster R-CNN：towards real-time object detection with region proposal networks［J］. IEEE Trans Pattern Anal Mach Intell，2017，39（6）：1137-1149.

18 Liu Z，Ning J，Cao Y，et al. Video Swin Transformer［J/OL］. https://doi.org/10.48550/arXiv.2106.13230.

19 Jiang T，Lu P，Zhang L，et al. RTMPose：Real-Time Multi-Person Pose Estimation based on MMPose［J/OL］. https://doi.org/10.48550/arXiv.2303.07399.

20 Nie Y，Nguyen NH，Sinthong P，et al. A Time Series is Worth 64 Words：Long-term Forecasting with Transformers［J/OL］. https://doi.org/10.48550/arXiv.2211.14730.

21 Devices for remote monitoring of Parkinson’s disease［EB/OL］. https://www.nice.org.uk/guidance/DG51.

（责任编辑：施晓亚，潘越）