International Journal of Computer Assisted Radiology and Surgery (2024) 19:831–840
https://doi.org/10.1007/s11548-023-03052-4

ORIGINAL ARTICLE

Pose-based tremor type and level analysis for Parkinson's disease from video

Haozheng Zhang[1] · Edmond S. L. Ho[2] · Francis Xiatian Zhang[1] · Silvia Del Din[3,4] · Hubert P. H. Shum[1]

Received: 14 April 2023 / Accepted: 20 December 2023 / Published online: 18 January 2024

Abstract

Purpose Current methods for diagnosis of PD rely on clinical examination. The accuracy of diagnosis ranges between 73 and 84%, and is influenced by the experience of the clinical assessor. Hence, an automatic, effective and interpretable supporting system for PD symptom identification would support clinicians in making more robust PD diagnostic decisions.

Methods We propose to analyze Parkinson's tremor (PT) to support the analysis of PD, since PT is one of the most typical symptoms of PD with broad generalizability. To realize the idea, we present SPA-PTA, a deep learning-based PT classification and severity estimation system that takes consumer-grade videos of front-facing humans as input. The core of the system is a novel attention module with a lightweight pyramidal channel-squeezing–fusion architecture that effectively extracts relevant PT information and filters noise. It enhances modeling performance while improving system interpretability.

Results We validate our system via individual-based leave-one-out cross-validation on two tasks: the PT classification task and the tremor severity rating estimation task. Our system presents a 91.3% accuracy and 80.0% F1-score in classifying PT with non-PT class, while providing a 76.4% accuracy and 76.7% F1-score in more complex multiclass tremor rating classification task.

Conclusion Our system offers a cost-effective PT classification and tremor severity estimation results as warning signs of PD for undiagnosed patients with PT symptoms. In addition, it provides a potential solution for supporting PD diagnosis in regions with limited clinical resources.

Keywords Parkinson's disease · Tremor type · Tremor rating · Deep learning · Channel attention

Introduction

Parkinson's disease (PD) is the second most common progressive neurological disorder, affecting an estimated 10 million people globally [1]. It is characterized by the loss of dopaminergic neurons within the substantia nigra region of the brain, resulting in motor dysfunction [2]. Existing PD diagnosis is mainly based on the clinical assessment of PD symptoms, medical history, l-dopa and dopamine responses [3]. The clinical diagnostic accuracy is approximately 73–84% [4] and may be affected by medical experts' subjective opinions and experiences. An automatic, efficient and interpretable PD assessment system would support clinicians in making more robust diagnostic decisions.

Recent research in PD diagnosis with machine learning using human-centric visual, audio and movement features has shown promising results. Models based on neuroimaging [5] and cerebrospinal fluid biomarkers [6] provide an accurate diagnosis but are costly and intrusive, making them unsuitable for large-scale pre-diagnosis. Non-intrusive methods with speech [7] are limited by their generalizability due to the significant difference in language and pronunciation for patients from different geographical areas. Although gait disturbance is not typically the primary symptom of early-onset PD [8, 9], over 70% of these patients exhibit at least one form of tremor [9]. Hence, identifying Parkinson's Tremor

✉ Hubert P. H. Shum
hubert.shum@durham.ac.uk

1 Department of Computer Science, Durham University, Durham, UK

2 School of Computing Science, University of Glasgow, Glasgow, UK

3 Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK

4 National Institute for Health and Care Research Newcastle Biomedical Research Centre, The Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK

(PT) is seen as a more generalizable approach for assisting in early PD diagnosis. To date, hand tremor-based studies mostly rely on wearable sensor data [10]. However, the use and setup of wearable technology may be time- and resource-consuming [10]. Video-based analysis with consumer-grade cameras is preferable as a more cost-effective solution without disrupting the natural behavior of the participants.

We propose a novel open-source[1] video-based deep learning system for PT classification and tremor severity estimation to assist the pre-diagnosis of PD with PT symptoms. We first extract the upper body human pose from videos as an effective feature for tremor analysis. We then design a graph neural network with a novel pyramidal channel-squeezing–fusion (PCSF) architecture that learns the attention by representing the joint-wise relevancy in a hierarchical manner. Such attention values allow interpretation of the features considered by the network for decision-making. Our solution outperforms existing ones in PT analysis, achieving 91.3% accuracy and 80.0% F1-score in PT classification, 76.4% accuracy and 76.7% F1-score in tremor rating classification.

Compared with our preliminary work [11] that only focuses on tremor-type classification, we have the following technical improvements: (1) adapting the system for tremor rating estimation; (2) supplementing our system with the Eulerian video magnification to enhance the subtle tremors for better feature extraction; (3) adding an examination with the Nyquist limits to test whether the input videos are suitable for tremor analysis; (4) improving pose extraction by employing the state-of-the-art AlphaPose algorithm and conducting comprehensive experiments to evaluate its performance improvement; (5) evaluating our system via a more challenging individual-based leave-one-out cross-validation to improve system robustness; and (6) conducting extra experiments with ablation studies and visualizations.

Method

Figure 1 shows the overview of our system. Its input is a set of videos showcasing a patient sitting in an upright posture, performing various actions such as keeping arms parallel to the ground. The human joint position features are extracted from the videos using AlphaPose [12], a state-of-the-art pose estimation algorithm. These features are then fed into the spatial pyramidal attention network for PT-type and level analysis (SPA-PTA).

Eulerian video magnification

We employ Eulerian video magnification (EVM) as a signal processing method [13] to enhance the subtle tremors and reduce noise and artifacts in the videos. This is motivated by previous research findings [11] that deep neural network models paid more attention to human wrists during PT classification, indicating that magnifying subtle hand and wrist motions can be beneficial for tremor feature learning. Before applying EVM, we checked the Nyquist limits [14] to examine whether our video frequency is valid for tremor analysis. Specifically, the video frame rate should be at least twice the highest frequency of tremor motions. As existing research [15] has shown that PT typically occurs between 3 and 7 Hz, our video with 30 Hz fulfills the requirement.

Pose extraction

We extract the 2D pose features from the EVM-processed videos by AlphaPose [12]. Compared to previous work using OpenPose [11], AlphaPose is superior as it demonstrates 25% improved pose estimation performance on average precision and average recall metrics in multiple datasets. We prefer 2D poses to 3D ones, as current 3D pose estimation techniques are less mature, and they generally introduce noise particularity in the depth dimension [16], making them less suitable for sensitive features like tremors. We use AlphaPose to estimate 17 COCO-format [12] body keypoints and extract (x, y, c) features, where (x, y) represent the 2D coordinate and c is a confidence score that reflects the estimation accuracy. Consistent with previous work [11], we utilize the top half of the body keypoints (shown in Fig. 5) for PT classification. It disregards less relevant lower body features to enhance model efficiency and reduce potential bias because of the observation that PT generally occurs on the upper body, specifically on the hands and arms [17]. In addition, we omit the head joints as the participants' faces are generally obscured in medical videos to preserve their privacy. Furthermore, we normalize the pose to mitigate bias resulting from inherent video differences. In order to mitigate global translations in the pose, we align the mean location of the neck and two hip joints as the global origin. Subsequently, all joint positions are expressed as relative values to this established origin.

Classification network

We propose the SPA-PTA for PT analysis by the PT classification task and an extended tremor severity estimation task. SPA-PTA is composed of two graph neural network (GNN) blocks with a spatial attention mechanism, along with a novel pyramidal channel-squeezing–fusion block designed to learn the joint-wise relevancy.

GNN block with spatial attention mechanism

We consider using graph neural networks (GNN) for PT analysis, which are effective in modeling relational data, unlike

[1] https://github.com/zhz95/SPA-TPA.

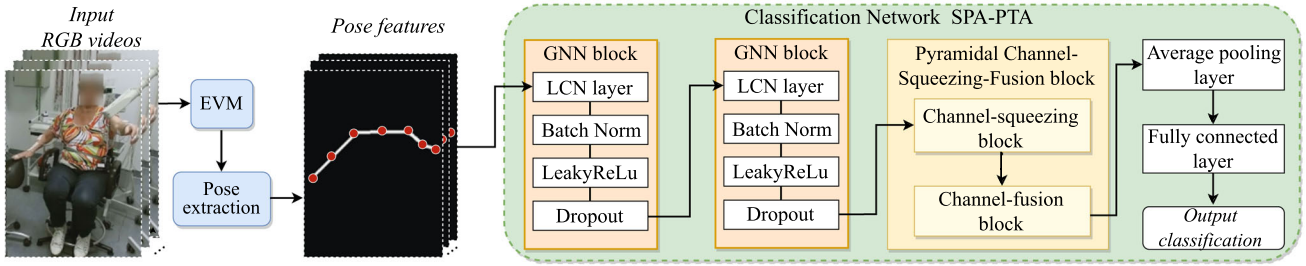

Fig. 1 Framework of our system: We use EVM to enhance the subtle tremors in the original videos, and then pass videos to the pose extraction process. We classify the extracted pose features by SPA-PTA with a novel PCSF design

images that are in a grid structure. In particular, human poses can be considered as a relational graph structure $G = (V, E)$ [18], with the nodes representing the joints and the edges representing the skeletal structure across time. Formally, $\{V = v_{\text{pq}}\}$ represents the set of joint positions, where v_{pq} is the p-th joint at the q -th frame. The set of edges, E, consists of (i) spatial edges connecting different joints in space and (ii) temporal edges connecting the same joint across consecutive frames.

We propose a spatial attention mechanism to enhance the performance of classification and improve the interpretability of our system. Specifically, it helps interpret the significant joints that the network identifies during classification by computing the attention weight of each joint per frame and its temporal aggregation. Moreover, it allows the system to learn the attention of the target joint by considering its relevancy with other joints. The fundamental expression is as $\mathbf{h_i} = \sigma\left(\sum_{j\in\mathcal{N}^i} \mathbf{W}_j^i \mathbf{x}_j \hat{a}_{ij}\right)$ where σ is an activation function, $\mathbf{W}_j^i$ is the learnable attention weight between the target node i and the related node j, $\hat{a}_{ij}$ is the corresponding element in the adjacency matrix, $\mathbf{x}_j$ is the input features of node j, $\mathcal{N}^i$ is the set of connected nodes for node i and $\mathbf{h}_i$ is the updated features of node i.

Pyramidal channel-squeezing–fusion block (PCSF)

We hypothesize that the relevancy between two joints depends on their proximity according to the skeletal structure. This aligns with information gain analysis [19], which proves that information gain diminishes exponentially as the node distance increases. Furthermore, clinical observation [20] suggests that PD patients typically experience PT on only one side of the upper body. Therefore, the information relevancy from one arm to another should be small.

To realize the hypothesis, we propose a novel lightweight PCSF that better models the relevancy of joints from their neighbors, thereby enhancing the network performance. As shown in Fig. 2, the output target node i's attention weight W^i is obtained from the joint-wise weights $\{W^i_{d_0}, \ldots, W^i_{d_{\max}}\}$ after the squeezing-and-fusion process, where d_n is the shortest distance between the target node i and the relevant node n, namely *Hop-n*. The visualization of information relevancy in Fig. 2 guides the squeezing ratio, such that our method overcomes the limitation of the GCN (graph convolutional network) [21] that each joint shares the same weight.

The channel-squeezing block We propose following squeezing operations to enhance the learning of PT-specific relevant information while filtering noise, based on our hypothesis motivated by [19, 20]. We distinguish nodes in different graph distance by defining *hop-0* node to be the self-node, *Hop-1,2* nodes to be the short-range nodes and *Hop-3,…,Hop-max* to be the long-range nodes. Suppose the node i is the target node, and the node j is the relevant node of i, then node j's output channel size is formulated by Eq. (1):

$$C_{\text{out},j} = \begin{cases} C_{\text{in}}, & |j-i| = 0, \\ pC_{\text{in}}, & 0 < |j-i| \le 2, \\ q^{|r-i|}C_{\text{in}}, & |j-i| > 2. \end{cases} \tag{1}$$

where p, q are channel-squeezing ratios for *Hop-1,2* nodes and *Hop-3,…, max* nodes, respectively. $p, q \in [0, 1]$ and $p \gg q$. $C_{\text{out},j}$ is the output channel size of node j. $|\cdot|$ denotes the graph distance between nodes.

The channel–fusion block To hierarchically combine the different range information of the target node i, we fuse the long-range features by f_l and fuse all features by f_a:

$$\mathbf{h_i} = f_a[\mathbf{h_{self}}, \mathbf{h_{short}}, f_l(\mathbf{h_{long,k}})]\mathbf{W^i} \tag{2}$$

where $\mathbf{h_{long,k}}$ is the feature of the long-range node k, $\mathbf{h_{short}}$ and $\mathbf{h_{self}}$ are features of short-range nodes and self-node, respectively, and $\mathbf{W^i}$ is the final weight matrix of target node i.

Implementation details

As depicted in Fig. 1, our network employs two GNN blocks with output channel sizes of 64 and 128, respectively. Each block contains an LCN (locally connected network [22])

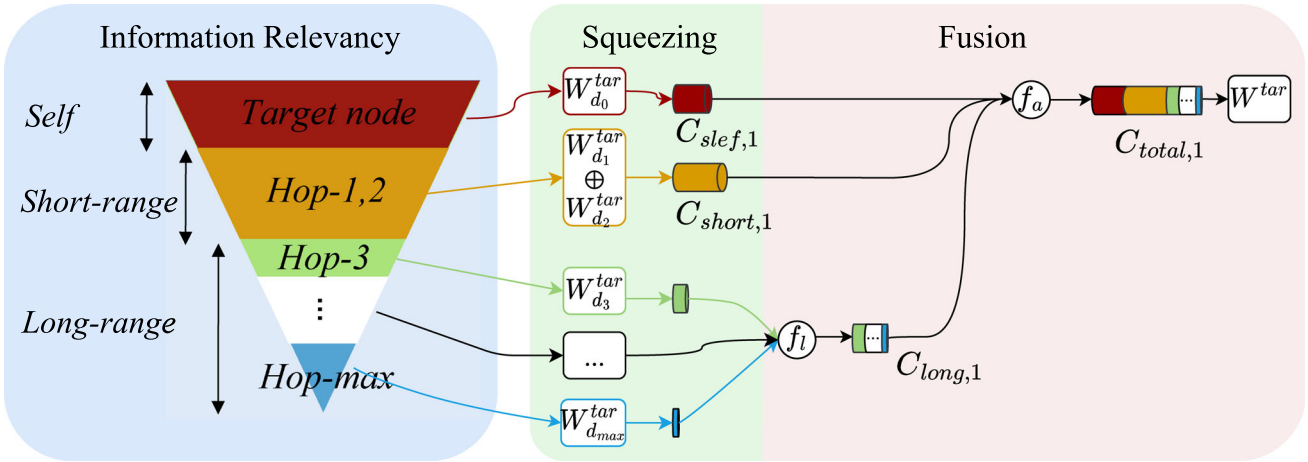

Fig. 2 Proposed pyramidal channel-squeezing–fusion architectures

layer, a batch normalization layer, a LeakyReLU layer with an alpha of 0.2 and a dropout layer with 0.2 rates. Following the two GNN blocks, we employ a PCSF block, a global average pooling layer and a fully connected layer. We adapt cross-entropy loss in binary classification. To address the class imbalance in multiclass classification, we use the focal loss [23] instead. Our optimizer of choice is Adam. The best performance of the PT binary classification task is achieved by a learning rate of 0.01 (decays by 0.1), a batch size of 8 and a dropout of 0.2, at 500 epochs.

Experiments

To assess the efficacy of our proposed method, we conducted validation testing on two separate evaluation examinations: the PT classification examination and the tremor rating estimation examination. We carried out our experiments using a Ubuntu 18.04 PC with an NVIDIA 3080. The GPU memory usage for training was minimal, averaging just 1.46 gigabytes. The training process for the TIM-TREMOR dataset took approximately ten hours for the PT classification task and twelve hours for the tremor rating estimation examination. They include the processes of EVM and extraction of human pose features from RGB videos. In terms of real-time application, the PT classification or tremor rating estimation of a 33-s video with 1000 frames only took around 48 s each, which is a feasible time for interactive diagnosis.

The dataset

We test our system using the TIM-TREMOR dataset [24], which is an open dataset consisting of 910 videos of 55 individuals performing 21 tasks. The videos are 18–112 s long. There are 572 videos depicting various forms of tremors, including 105 for parkinsonian tremor (PT), 182 for essential tremor (ET), 88 for functional tremor (FT) and 197 for dystonic tremor (DT). An additional 60 videos (NT) were recorded without convincing tremors during the assessment. The test 278 videos have inconclusive tremor classification results and have been labeled as "Other." For the tremor rating labels, eight levels from level 0 to 7 are assigned to the individual's left and right hands, evaluated by Bain and Findley Tremor Clinical Rating Scale [25]. To ensure that there is only one label per video and preserve the characteristics of the video, we combine the labels for individual left and right hands by taking the maximum value of both hands.

Setup

We eliminate inconsistent videos to minimize data noise, specifically videos that only capture motion tasks for a limited number of participants. For the tremor-type classification task only, we remove the videos with uncertain tremor-type labels of "other." Next, each video is clipped into 100-frame samples, and the number of samples is determined by the duration of the consecutive frames in which the participant was visible and not obscured. Each sample was assigned the label of the original video and treated as an individual sample. We use a voting system to obtain the video-level classification results, which increases the system's robustness and augments the training sample size [26]. We evaluate our proposed system through individual-based leave-one-out cross-validation. It means each subclip for a single individual is used for testing and excluded from the training set for each iteration. The subclips for each individual are never separated by the training or testing set. The total number of leave-one-out cross-validations are 39 and 55 for tremor-type classification and tremor rating estimation, respectively.

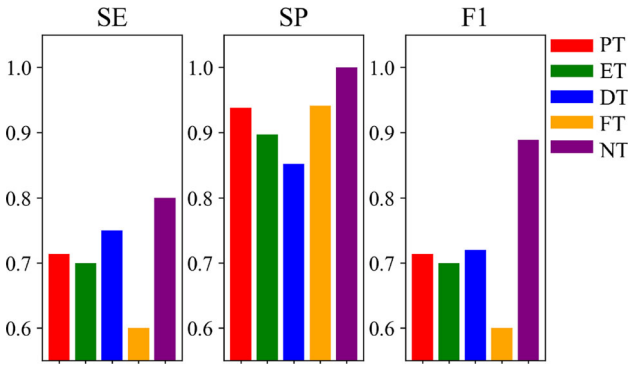

Fig. 3 Per-class multiclass tremor-type classification results

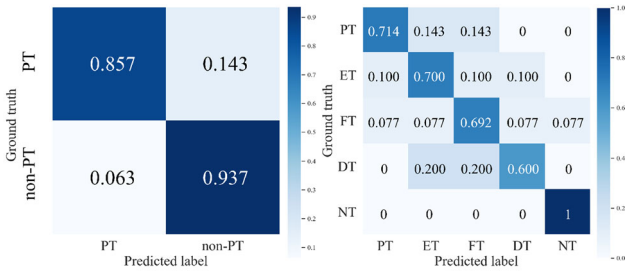

Fig. 4 Confusion matrices for PT classifications: (Left) binary; (Right) multiclass

Evaluation metrics

We report the mean values calculated among all leave-one-out cross-validations with the following metrics: accuracy (AC), sensitivity (SE), specificity (SP) and F1-score for the binary classification; AC, macro-average F1-score, SE and SP for the multiclass classification.

Tremor-type classification

For this experiment, we first evaluate our system on the binary classification that distinguishes PT labels from non-PT labels, and achieve 91.3% accuracy and 80.0% F1-score. In addition, we validate our method on a more complex multiclass classification task for classifying five types of tremors (PT, ET, DT, FT and NT). Our final system's per-class tremor-type multiclass classification performance is shown in Fig. 3. It shows a fairly balanced performance on classifying PT, ET, DT and NT, while FT has a lower SE and F1-score, which may be caused by the smallest number of samples in this class. Moreover, the corresponding confusion matrices of the two tasks are displayed in Fig. 4.

Comparison with baseline methods

As this paper is the first work that provides the individual-level evaluation results, we implemented the following video-based PT classification baselines to evaluate the effectiveness of our system: (i) ST-GCN [18]: a spatial–temporal GCNs for human pose data classification; (ii) CNN with 1D convolutional layers (CNN-Conv1D) [27]; (iii) decision tree (DT); and (iv) support vector machine (SVM) [27]. Note that all baseline methods apply the same EVM and pose extraction design. The results of our proposed SPA-PTA and baselines are summarized in Table 1.

The binary classification result shows that our full system consistently outperforms all other methods in all evaluation metrics. Our AC, SE, SP and F1 achieve over 80% on leave-one-out cross-validation, demonstrating the effectiveness and stability of our system in this binary classification task. It is noticeable that our system performs better with only spatial convolution instead of a deeper spatial–temporal convolution design like ST-GCN [18]. The outcome supports that the suggested PCSF block effectively enhances classification reliability and reduces the risk of overfitting in small datasets.

While the full system is initially designed for binary classification, it presents effectiveness and generalizability in the multiclass classification task, surpassing existing methods. A small difference between AC, SE and SP shows that our system performs consistently and effectively at identifying the positive samples and excluding the negative ones. The high macro-average SP exhibited trustworthy effectiveness in correctly recognizing individuals who have a specific type of tremor without wrongly recognizing it as other types of tremor.

Ablation studies

We conduct an ablation analysis to assess the effectiveness of the EVM, PCSF block and the entire attention module. From the lower parts of Table 1, the positive effect of the PCSF block and attention module can be illustrated by the decrease in metrics when either the PCSF block or the entire attention module is removed in the two classification tasks. Also, we find that the basic GNN architecture without attention performs better than the CNN-Con1D model for both classification tasks. It highlights the efficacy of learning human pose features in the graph domain as opposed to the Euclidean domain. Besides, the variant of "ours without attention" performs slightly better than "ours without attention and EVM preprocessing," indicating that the use of EVM could effectively enhance tremors.

Model interpretation

We present the visualization for the average attention value of each body keypoint in Fig. 5a. It is interpreted as the importance level our system considers during the classification process. Our analysis reveals that the attention value

Table 1 Comparisons on the tremor-type classification task

Method	Binary classification				Multiclass classification			
	AC	SE	SP	F1	AC	SE	SP	F1
ST-GCN [18]	84.6	71.4	87.5	62.5	64.1	64.8	90.7	64.1
CNN-Conv1D	76.9	57.1	81.3	47.1	56.4	54.2	88.3	53.1
Decision Tree	69.2	57.1	71.9	40.0	51.3	49.4	87.6	36.7
SVM [27]	64.1	57.1	65.6	36.4	46.2	44.6	86.2	44.3
Ours-full	**92.3**	**85.7**	**93.8**	**80.0**	**71.8**	**71.3**	**92.5**	**72.5**
w/o PCSF	87.2	85.7	87.5	70.6	66.7	67.6	91.4	66.7
w/o Attention	82.1	71.4	84.4	58.8	61.5	63.1	90.0	62.4
w/o Attention & EVM	79.5	71.4	81.3	55.6	59.0	59.1	89.5	58.5

Best performance results are highlighted in bold

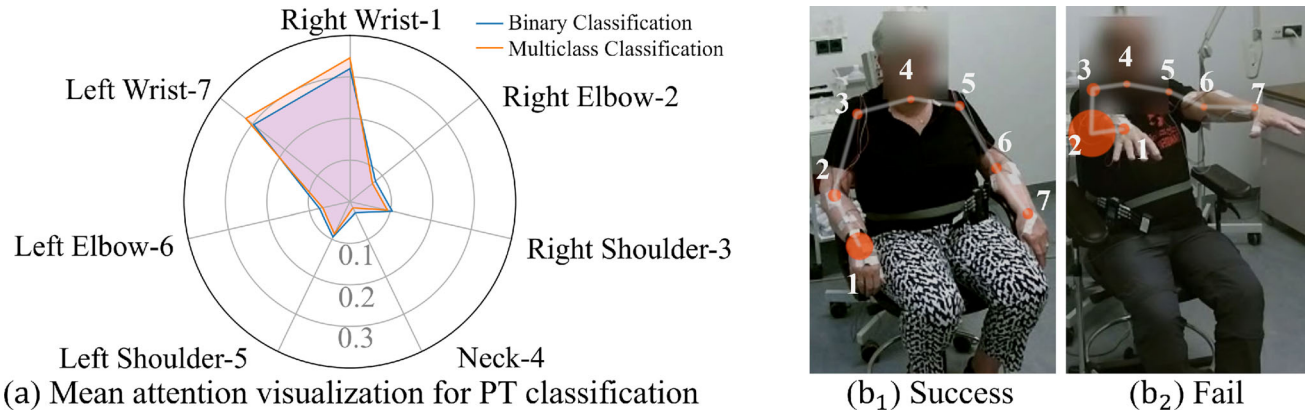

Fig. 5 (**a**) Average skeleton joints attention across all cross-validations in the PT classification experiment. (**b**) Attention visualization at a (b_1) successfully classified frame, and (b_2) unsuccessful classified frame. The joint labels in (b) correspond to (a)

is significantly highest on the "Right Wrist" and "Left Wrist," which suggests that our system prioritizes the wrists' movements during the task performance. Furthermore, the value associated with the "Neck" is significantly lower than other keypoints. It may be explained by the fact that the participants remained seated during the video recording, resulting in a minimal global variance of the neck joint throughout the experiment.

Tremor rating estimation

For this experiment, we train SPA-PTA with different tremor rating labels without further implementation (e.g., converting the classification layer to a regression layer) to validate our system performance in the tremor rating estimation task. Since the data with tremor ratings 4 and above is insufficient for training via leave-one-out cross-validation (i.e., only 5 individuals out of 55), we validate our system on two different classification settings: (1) classifying ratings [1, 2, 3] and (2) classifying ratings [1, 2, 3+]. The latter is generally a more challenging task since the imbalanced data of the "3+" rating brings bias compared to the former, which does not contain such data (Figs. 6, 7).

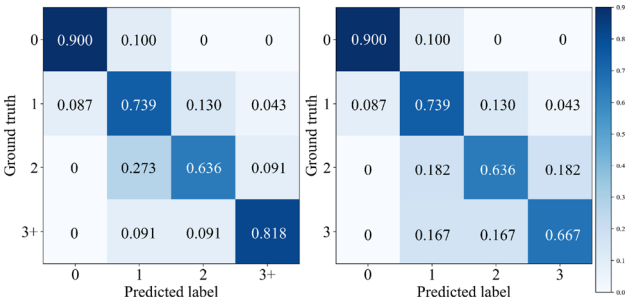

Fig. 6 Confusion matrices for tremor rating estimation: (Left) [1, 2, 3+]; (Right) [1, 2, 3]

Comparison with Baseline methods

We compare our SPA-PTA to the same baselines in the tremor-type classification task as shown in Table 2. SPA-PTA significantly outperforms the baselines by achieving a macro-average AC of 76.4%, SE of 77.3%, SP of 91.6% and F1-score of 76.7%. An interesting finding is that the machine learning-based method decision tree achieves similar performance to two deep learning-based baselines (i.e., ST-GCN and CNN-Conv1D). It may inform us to tackle the

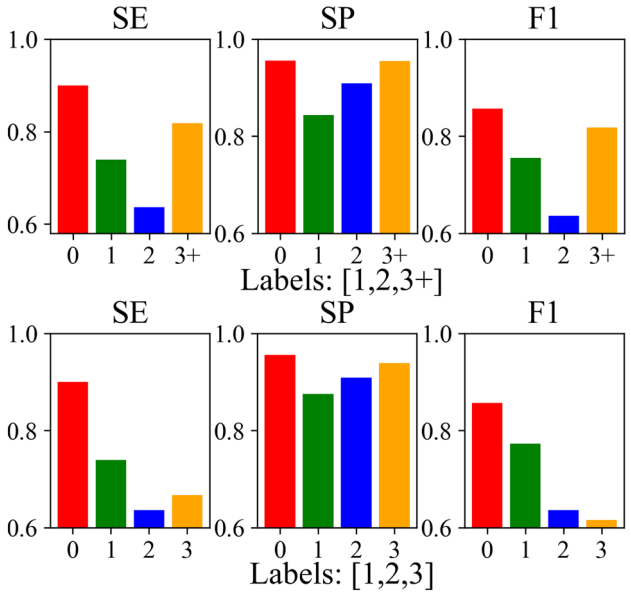

Fig. 7 Per-class tremor rating estimation results

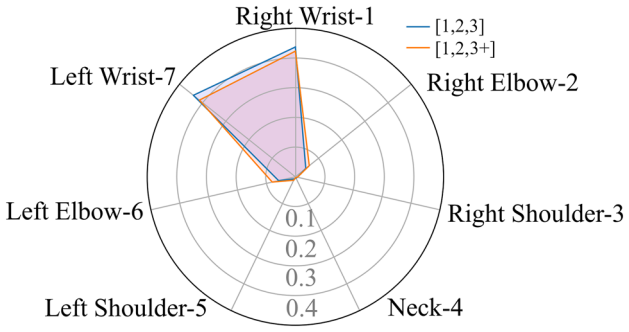

Fig. 8 Average skeleton joint attention across all cross-validations in tremor rating estimation task

challenge of improving the deep learning models in a relatively small dataset. In addition, although our current model does not show strong robustness in the tremor rating estimation task, the ablation studies from the rows of "Ours" in Table 2 still demonstrate the effectiveness of our PCSF layer and the attention mechanism design. It shows the potential of improving our model and system performance with a more specific architecture design with a more extensive dataset.

Ablation studies

Consistent results at the bottom of Table 2 from the same ablation design as for the PT classification task validate the effectiveness of each system component.

Model interpretation

We similarly visualize the average skeleton joints attention across all cross-validation sets in Fig. 8. Two different data preprocessing approaches provide similar attention results, while the weights obtained by grouping [1, 2, 3] slightly more contribute to "Right Wrist" and "Left Wrist." This may be due to the increased proportion of low tremor rating videos in this approach compared to grouping [1, 2, 3+]. In addition, we notice that the attention weight distribution of the tremor rating estimation examination is similar to that of the PT classification examination, while the former aggregates more attention on the "Right Wrist" and "Left Wrist" than other joints.

Pose estimation evaluation

To evaluate the effectiveness of AlphaPose and quantify the pose estimation error, we conduct the following experiments:

Quantitative comparison with ground truth data

To quantify the pose estimation error from different methods, we employ the Lagrangian hand-tremor frequency estimation method [24] to compare MAE (mean absolute error) of the hand-tremor frequencies estimated by AlphaPose and conventional OpenPose features [11] with ground truth (GT) frequency obtained from accelerometer data. As suggested in [24], tremor frequency calculated from reliable estimated

Table 2 Comparisons on the tremor rating task

Classification labels	[1, 2, 3+]				[1, 2, 3]			
Method	AC	SE	SP	F1	AC	SE	SP	F1
ST-GCN [18]	67.3	68.1	89.0	66.5	68.0	67.7	90.5	65.7
CNN-Conv1D	60.0	59.8	86.5	58.7	60.0	60.5	87.9	58.3
Decision Tree	54.5	55.3	85.2	54.6	52.0	53.0	86.0	51.3
SVM [27]	49.1	41.1	81.5	43.8	48.0	49.5	85.2	47.1
Ours-full	**76.4**	**77.3**	**91.6**	**76.7**	**74.0**	**73.5**	**92.0**	**72.0**
w/o PCSF	70.9	71.5	89.7	70.7	70.0	68.6	90.5	68.2
w/o Attention	65.5	65.6	88.2	64.8	66.0	65.2	89.5	63.9
w/o Attention & EVM	63.6	64.8	87.6	63.3	64.0	64.1	88.9	62.5

Best performance results are highlighted in bold

Table 3 MAE comparison between AlphaPose features and OpenPose on the top-10 best-performing tasks

Task	AlphaPose	OpenPose
Rest	**0.812**	0.881
Rest in supination	**0.834**	0.930
2 Hz higher	**0.605**	0.635
2 Hz lower	**0.617**	0.622
Counting	**0.729**	0.790
Finger tapping	**0.576**	0.687
Playing piano	**0.752**	0.906
Months backward	**0.814**	0.838
Top top	**0.786**	0.823
Thumbs up	**0.844**	0.960
Average MAE	**0.737**	0.807

Better performance with lower MAE is in bold

pose features should be close to (i.e., ideally within 1 HZ difference) the GT accelerometer data frequency. The MAE from Table 3 indicates that AlphaPose consistently outperforms OpenPose on all listed tasks.

Qualitative pose visualization and comparison

The visualizations in Fig. 9 and the reference video images in Fig. 10 show that AlphaPose outperforms OpenPose in estimating joint positions. This is supported by the smoother trajectory lines of AlphaPose, which are depicted by the transparent colored lines. Figures 1, 2, 3, 4, 5 and 9 demonstrate AlphaPose's ability to track joint movement fluidly. Specifically, in Fig. 5, AlphaPose demonstrates a hand trajectory that aligns more closely with the anticipated tremor pattern, which contrasts with OpenPose's intermittent jumping trajectory. This consistency suggests that AlphaPose may be more reliable for tasks related to PT classification. Furthermore, on the patient's right side, particularly in Figs. 1 and 2, AlphaPose yields more consistent and stable outcomes, reflecting the patient's condition of resting with observable tremors only in the left hand, as corroborated by Fig. 10. Finally, the neck joint position of OpenPose is estimated by the mean point of both shoulders, which is less accurate than the estimated neck joint position of AlphaPose [12].

Classification performance comparison

We compare the effectiveness of AlphaPose and OpenPose by evaluating their impacts on the system classification performance. Table 4 demonstrates that using AlphaPose features results in a remarkable and consistent improvement over OpenPose features of approximately $1-3\%$ across the classification tasks except for the binary tremor-type classification. These results highlight the precision of AlphaPose in delivering better pose-based features for classification tasks.

In this study, we utilize the pre-trained AlphaPose model, opting not to retrain it due to the absence of GT 2D pose position annotations within our dataset. The robust generalization capability of the pre-trained AlphaPose model, as evidenced

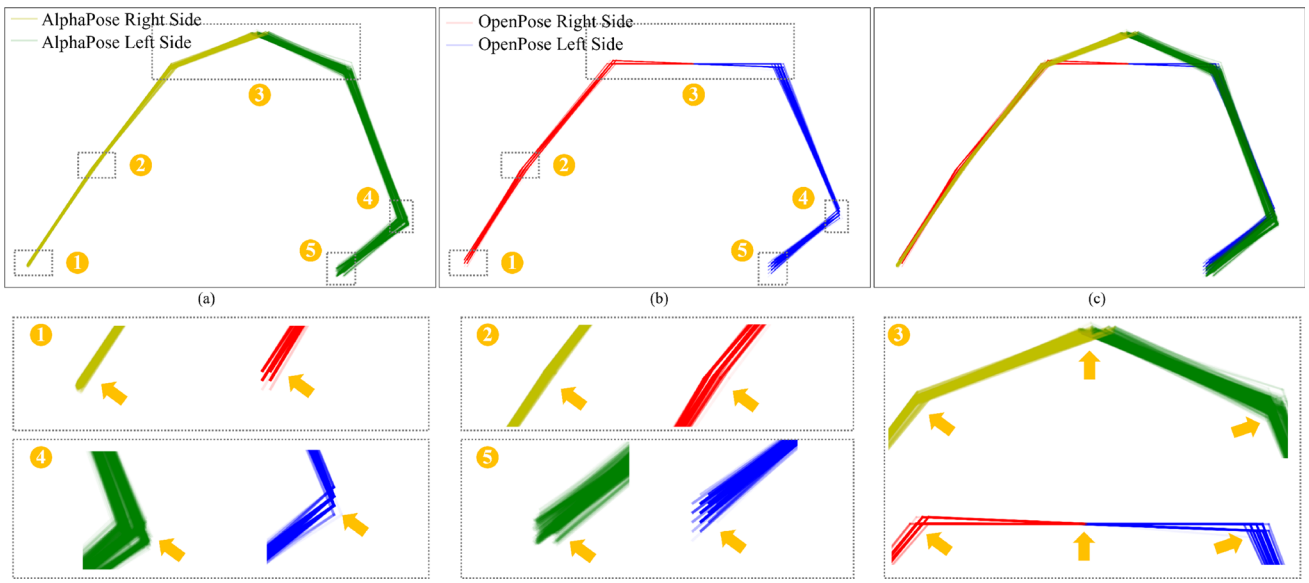

Fig. 9 Estimated pose comparison between AlphaPose and OpenPose for a sitting and resting PD patient with clinically identified PT on the left side of the body. (**a**)–(**c**) are the estimated poses of an example video from AlphaPose, OpenPose and both, respectively. Each colored line with 0.05 transparency represents the connection between joints estimated in each frame. Numbers 1 to 5 correspond to specific joints' local scaling for intuitive comparison. The raw video frames are referenced in Fig. 10

Fig. 10 Raw videos referenced in Fig. 9 consist of consecutive images captured at intervals of 5 frames, approximately every 0.167 s. The lower right image is an aggregation of five transparent hand images, where the green dot shows the estimated trajectory of the left wrist joint during tremor

Table 4 Comparisons on the influence of classification performance between AlphaPose and OpenPose

Method	AC	SE	SP	F1	AC	SE	SP	F1
	Tremor-type Binary				Tremor-type Multiclass			
AlphaPose	**92.3**	**85.7**	**93.8**	**80.0**	**71.8**	**71.3**	**92.5**	**72.5**
OpenPose	**92.3**	**85.7**	**93.8**	**80.0**	69.2	69.3	92.0	70.2
	Tremor level [1,2,3+]				Tremor level [1,2,3]			
AlphaPose	**76.4**	**77.3**	**91.6**	**76.7**	**74.0**	**73.5**	**92.0**	**72.0**
OpenPose	72.7	74.0	90.4	73.6	72.0	72.5	90.4	70.1

Best performance results are highlighted in bold

by its superior performance across multiple diverse and complex benchmark datasets [12], affirms its suitability for our task. In the future, we are interested in comparing the performance between pre-trained and tremor-specific pose estimation models. This will entail the collection of the necessary GT data to train a model adept at detecting the subtle nuances characteristic of tremor movement patterns.

Conclusion and discussion

Our method effectively identifies PT in PD patients from consumer-grade videos. The validity of our proposed system on both PT classification and tremor severity estimation tasks demonstrates that our method is extensible in PT-related analysis. Our non-intrusive system only relies on consumer-grade videos as input, so it offers a potentially cost-effective solution for supporting the pre-diagnosis of PD in regions with inadequate medical resources. This work could also be used for remote PD supplementary assessment in special situations to reduce the stress of the healthcare system (e.g., the COVID-19 pandemic). Moreover, our system demonstrates the potential to automatically monitor PT symptoms during daily activities to support PD pre-diagnosis.

Our findings about PT analysis are preliminary, and the limited number of people with PT and the limited range of tremor levels included in this work may affect the generalizability of the results. One of our future directions is to evaluate our models using data collected from a larger and more diverse group of Parkinson's disease patients, covering a more balanced tremor-type distribution and a wider range of tremor severity ratings. Upscaling the study is crucial for developing more robust models and for enhancing the overall applicability and validity of the framework we have presented. In addition, annotating the dataset based on PT severity estimation performance by different scales, such as the MDS-UPDRS3, by experienced raters will enable us to improve the robustness of our model in the future. Moreover, our current system performance is still challenged by pose estimation algorithm error, such as depicted in blue Fig. 5b. The attention of our system is incorrectly influenced by the inaccurate position detection of the right elbow and blurred right shoulder joints.

Funding H. Shum received support from the EPSRC NorthFutures project (Ref: EP/X031012/1). S. Del Din has received support from Innovative Medicines Initiative 2 Joint Undertaking (Ref: 820820 Mobilise-D, 853981 IDEA-FAST), NIHR Newcastle, Newcastle upon Tyne Hospitals NHS Foundation Trust, Cumbria Northumberland and Tyne and Wear NHS Foundation Trust.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval Approval of the TIM-TREMOR dataset was obtained from the University Leiden University Medical Center ethics committee. The procedures used in this study adhere to the tenets of the Declaration of Helsinki.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Chopade P, Chopade N, Zhao Z, Mitragotri S, Liao R, Chandran Suja V (2023) Alzheimer's and Parkinson's disease therapies in the clinic. Bioeng Transl Med 8(1):10367
2. Mhyre TR, Boyd JT, Hamill RW (2012) Parkinson's disease. Subcellular Biochem 65:389–455
3. Mostafa SA, Mustapha A, Mohammed MA, Hamed RI, Arunkumar N, Ghani M, Jaber MM, Khaleefah SH (2019) Examining multiple feature evaluation and classification methods for improving the diagnosis of Parkinson's disease. Cognit Syst Res 54:90–99
4. Rizzo G, Copetti M, Arcuti S, Martino D, Fontana A, Logroscino G (2016) Accuracy of clinical diagnosis of Parkinson disease: a systematic review and meta-analysis. Neurology 86(6):566–576
5. Zhang L, Wang M, Liu M, Zhang D (2020) A survey on deep learning for neuroimaging-based brain disorder analysis. Front Neurosci 14:779
6. Wang W, Lee J, Harrou F, Sun Y (2020) Early detection of Parkinson's disease using deep learning and machine learning. IEEE Access 8:147635–147646
7. Vásquez-Correa JC, Arias-Vergara T, Orozco-Arroyave JR, Eskofier B, Klucken J, Nöth E (2018) Multimodal assessment of Parkinson's disease: a deep learning approach. IEEE J Biomed Health Inform 23(4):1618–1630
8. Hausdorff JM (2009) Gait dynamics in Parkinson's disease: common and distinct behavior among stride length, gait variability, and fractal-like scaling. Chaos 19(2):026113
9. Rizek P, Kumar N, Jog MS (2016) An update on the diagnosis and treatment of Parkinson disease. CMAJ 188(16):1157–1165
10. Hssayeni MD, Jimenez-Shahed J, Burack M, Ghoraani B (2019) Wearable sensors for estimation of parkinsonian tremor severity during free body movements. Sensors 19(19):4215
11. Zhang H, Ho ESL, Zhang X, Shum HPH (2022) Pose-based tremor classification for Parkinson's disease diagnosis from video. In: MICCAI
12. Fang H, Li J, Tang H, Xu C, Zhu H, Xiu Y, Li Y-L, Lu C (2022) Alphapose: whole-body regional multi-person pose estimation and tracking in real-time. IEEE Trans Pattern Anal Mach Intell 45:1–17
13. Liu W, Lin X, Chen X, Wang Q, Wang X, Yang B, Cai N, Chen R, Chen G, Lin Y (2023) Vision-based estimation of MDS-UPDRS scores for quantifying Parkinson's disease tremor severity. MIA 85:102754
14. Nyquist H (1928) Certain topics in telegraph transmission theory. Trans Am Inst Electr Eng 47(2):617–644
15. Delval A, Rambour M, Tard C, Dujardin K, Devos D, Bleuse S, Defebvre L, Moreau C (2016) Freezing/festination during motor tasks in early-stage Parkinson's disease: a prospective study. Mov Disord 31(12):1837–1845
16. Wang J, Yan S, Xiong Y, Lin D (2020) Motion guided 3d pose estimation from videos. In: ECCV, pp 764–780
17. Sveinbjornsdottir S (2016) The clinical symptoms of Parkinson's disease. J Neurochem 139:318–324
18. Yan S, Xiong Y, Lin D (2018) Spatial temporal graph convolutional networks for skeleton-based action recognition. In: AAAI, vol 32
19. Zhang W, Wang Y, You Z, Cao M, Huang P, Shan J, Yang Z, CUI B (2022) Information gain propagation: a new way to graph active learning with soft labels. In: ICLR
20. Fahn S (2003) Description of Parkinson's disease as a clinical syndrome. Ann N Y Acad Sci 991(1):1–14
21. Kipf TN, Welling M (2017) Semi-supervised classification with graph convolutional networks. In: ICLR
22. Ci H, Ma X, Wang C, Wang Y (2020) Locally connected network for monocular 3d human pose estimation. IEEE Trans Pattern Anal Mach Intel 44(3):1429–1442
23. Lin T, Goyal P, Girshick R, He K, Dollár P (2017) Focal loss for dense object detection. In: ICCV, pp 2980–2988
24. Pintea SL, Zheng J, Li X, Bank P, van Hilten JJ, van Gemert JC (2018) Hand-tremor frequency estimation in videos. In: ECCV
25. Bain PG, Findley LJ, Atchison P, Behari M, Vidailhet M, Gresty M, Rothwell JC, Thompson PD, Marsden CD (1993) Assessing tremor severity. J Neurol Neurosurg Psychiatry 56(8):868–873
26. Lu M, Zhao Q, Poston KL, Sullivan EV, Pfefferbaum A, Shahid M, Katz M, Kouhsari LM, Schulman K, Milstein A (2021) Quantifying Parkinson's disease motor severity under uncertainty using MDS-UPDRS videos. Med Image Anal 73:102179
27. Wang X, Garg S, Tran SN, Bai Q, Alty J (2021) Hand tremor detection in videos with cluttered background using neural network based approaches. Health Inf Sci Syst 9(30):1–14