Pose-based Tremor Classification for Parkinson's Disease Diagnosis from Video

Haozheng Zhang[1 [0000−0003−1312−4566]], Edmond S. L. Ho[2 [0000−0001−5862−106X]], Xiatian Zhang[1 [0000−0003−0228−6359]], and Hubert P. H. Shum[(✉)1 [0000−0001−5651−6039]]

[1] Durham University, UK
[2] Northumbria Univeristy, UK
{haozheng.zhang,xiatian.zhang,hubert.shum}@durham.ac.uk
e.ho@northumbria.ac.uk

Abstract. Parkinson's disease (PD) is a progressive neurodegenerative disorder that results in a variety of motor dysfunction symptoms, including tremors, bradykinesia, rigidity and postural instability. The diagnosis of PD mainly relies on clinical experience rather than a definite medical test, and the diagnostic accuracy is only about 73-84% since it is challenged by the subjective opinions or experiences of different medical experts. Therefore, an efficient and interpretable automatic PD diagnosis system is valuable for supporting clinicians with more robust diagnostic decision-making. To this end, we propose to classify Parkinson's tremor since it is one of the most predominant symptoms of PD with strong generalizability. Different from other computer-aided time and resource-consuming Parkinson's Tremor (PT) classification systems that rely on wearable sensors, we propose SPAPNet, which only requires consumer-grade non-intrusive video recording of camera-facing human movements as input to provide undiagnosed patients with low-cost PT classification results as a PD warning sign. For the first time, we propose to use a novel attention module with a lightweight pyramidal channel-squeezing-fusion architecture to extract relevant PT information and filter the noise efficiently. This design aids in improving both classification performance and system interpretability. Experimental results show that our system outperforms state-of-the-arts by achieving a balanced accuracy of 90.9% and an F1-score of 90.6% in classifying PT with the non-PT class.

Keywords: Parkinson's diagnosis · Tremor analysis · Graph Neural Network · Attention Mechanism · Deep Learning.

1 Introduction

Parkinson's disease (PD) is a progressive neurodegenerative disorder characterized by a variety of life-changing motor dysfunction symptoms, including tremor, bradykinesia (slow of movement), rigidity (limb stiffness), impaired balance and gait [14]. According to pathological studies, the motor deficits of PD are mainly caused by the loss of dopamine due to the degeneration of dopamine neurons

in patients [20]. As the second most common neurological disorder, the diagnosis of PD mainly relies on clinical criteria based on the parkinsonian symptoms (e.g., tremor, bradykinesia), medical history, and l-dopa or dopamine response [10,30,21]. However, the clinical diagnostic accuracy of PD is only about 73-84% [25] since the diagnostic performance is challenged by the subjective opinions or experiences of different medical experts [19]. Therefore, an efficient and interpretable automatic PD diagnosis system is valuable for supporting clinicians with more robust diagnostic decision-making.

Recent machine learning and deep learning-based methods achieved impressive performance in PD diagnosis by analyzing the neuroimaging, cerebrospinal fluid, speech signals, gait pattern [1], and hand tremors. Although neuroimagings [33] or cerebrospinal fluid [29] based models perform well, they face a problem of high cost and intrusive. As for the non-intrusive methods, current speech-based models [7] are limited by their generalizability, as the language and pronunciation habits of people in different regions and countries vary significantly. Several studies [11,24] indicate that gait disturbance is less likely to be the main symptom in patients with early-onset PD, but more than 70% of those patients present at least one type of tremors [3,22,24]. Hence we believe that detecting PD by diagnosing Parkinson's Tremor (PT) is a more generalizable approach compared with other methods. Conventional hand tremors-based studies [12] achieve promising performance by using a deep learning network on wearable sensors data to detect PD. However, using wearable sensors is still time and resource-consuming [12], and requires careful synchronization of data captured from different sensors.

For the first time, we propose a graph neural network for diagnosing PD by PT classification as it effectively learns the spatial relationship between body joints from graph-structured data. Inspired by the information gain analysis [8] and the clinician observation [9] that PT usually occurs only on one side of the early stage PD patient's upper body, we propose a novel attention module with a lightweight pyramidal channel-squeezing-fusion architecture to capture the self, short and long-range joint information specific to PT and filter noise. This design aids in improving both classification performance and system interpretability. Our system only requires consumer-grade non-intrusive video recordings and outperforms state-of-the-arts by achieving a balanced accuracy of 90.9% and an F1-score of 90.6% in classifying PT with non-PT class. Our work demonstrates the effectiveness and efficiency of computer-assisted technologies in supporting the diagnosis of PD non-intrusively, and provides a PT classification warning sign for supporting the diagnosis of PD in the resource-limited regions where the clinical resources are not abundant. Our source code is available at: https://github.com/mattz10966/SPAPNet.

2 Method

As shown in Fig. 1, the input consists of video recordings of each participant sitting in a chair in a normal upright position with various poses (e.g., tapping

with the contralateral hand in the rhythm). We extract the human joint position features from the RGB video by OpenPose algorithm [4]. These human joint position features are passed to the Spatial Pyramidal Attention Parkinson's tremor classification Network (SPAPNet) for diagnosis.

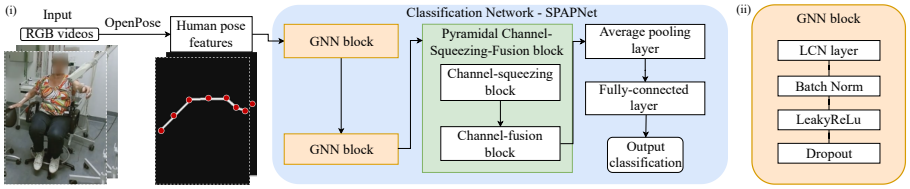

Fig. 1. (i) The overview of our proposed framework. (ii) The design of each GNN block.

2.1 Pose Extraction

We first extract 2D skeleton features from the video sequences. Each frame is fed to OpenPose [4] due to its robust and efficient performance in detecting the 2D joint landmarks for people in normal upright positions. We do not estimate the 3D human pose as in [16], since the state-of-the-art 3D pose estimation methods still introduce noise while processing the 2D information to 3D [5,18,27], which is not suitable for sensitive features like the tremor. We extract 18 OpenPose-skeleton format [4] landmarks with 2D coordinate (x, y) and a confidence score c indicating the estimation reliability by the OpenPose, but only use the seven upper body landmarks (seen in Fig. 3) for PT classification, because PT usually tends to occur on the upper body, especially the hands and arms [26]. This approach eliminates less relevant features to help reduce model bias and improve efficiency. In addition, we do not include the head joint considering the participant's privacy, since the face is generally occluded in the medical video. We implement normalization to reduce the bias from the underlying differences between the videos to tackle overfitting risk. To remove the participants' global translation, we center the participant's pose per frame by aligning the center of the triangle of the neck and two hip joints as the global origin. Then, we represent all joints as a relative value to the global origin.

2.2 Classification Network

We propose a *Spatial Pyramidal Attention Parkinson's tremor classification Network (SPAPNet)* for PT diagnosis. The proposed SPAPNet consists of a graph neural network with the spatial attention mechanism and a novel pyramidal channel-squeezing-fusion block to enhance the attention mechanism.

Graph Neural Network with Spatial Attention Mechanism:

Graph Neural Network (GNN): We propose to use the graph neural network to diagnose PD by classifying PT, since it effectively learns the spatial relationship between human joints from graph-structured data (e.g., human poses). To this end, we follow [31] to apply a pose graph $G = (V, E)$ aligned with the human skeletal graph to structure human pose data in the graph domain. In this graph, $\{V = v_{pq}\}$ denotes the joints positions, where v_{pq} represents the p-th joint at q-th frame. The edge set E includes: (1) the intra-skeleton connection each frame designed by the natural connections of human joints. (2) the inter-frame connections which connect the joints in consecutive frames.

Spatial Attention Mechanism: To improve the PT classification performance and interpret system the by human joints' importance, we propose using the spatial attention mechanism. Specifically, it interprets the important joints that the network considers in PT classification at each frame and video by attention weights and the temporal aggregation of the attention weights, respectively.

We adopt the locally connected network (LCN) [6] to learn joint i's attention weight from its relationship between other joints. This method overcomes the representation power limitation that different joints share the same weight set in the vanilla graph convolutional network (GCN) [13]. In addition, it enables the system to learn joint i's attention from its relationship between other joints. The basic formulation is as follows:

$$\mathbf{h_i} = \sigma\left(\sum_{j \in \mathcal{N}^i} \mathbf{W}_j^i \mathbf{x}_j \hat{a}_{ij}\right) \tag{1}$$

where $\mathbf{W}_j^i$ is the learnable attention weight between the target joint i and the related joint 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, σ is an activation function, and $\mathbf{h}_i$ is the updated features of node i.

Pyramidal Channel-Squeezing-Fusion Block (PCSF): As an extension of the spatial attention module, we propose a novel lightweight inverted pyramid architecture consisting of a *channel-squeezing block* and a *channel-fusion block* to extract relevant PT information and filter noise. This is motivated by two findings: (i) Information Gain analysis [8] shows that the information gain decreases exponentially with increasing distance between graph nodes; (ii) clinical observation [9] shows that PT usually occurs only on one side of the PD patient's upper body, such that the information relevancy between two arms should be reduced. Our proposed design does not require learnable parameters, such that it prevents overfitting problems. As illustrated in Fig. 2, we introduce the proposed PCSF by comparing it with the vanilla weight-sharing strategy in GCN [13]. In PCSF, the final attention weight for joint-1 is learned from the information between the target joint 1 and the relevant joints 2,3,...,7 after a series of channel squeezing and fusion operations. Conversely, the vanilla weight-sharing mechanism can not learn from the joint-wise relevancy since all joints share the same set of weights.

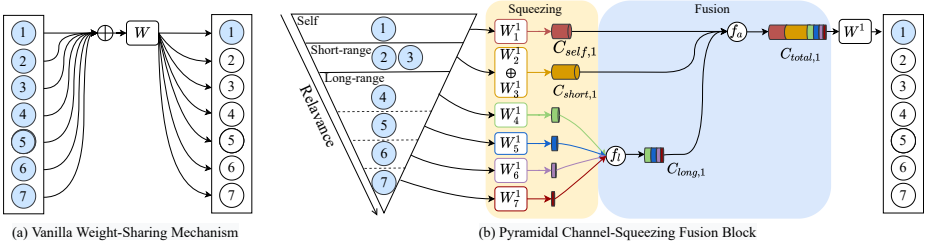

Fig. 2. The architectures of (a) Vanilla weight-sharing mechanism in GCN [13], (b) Proposed Pyramidal Channel-Squeezing-Fusion (PCSF) mechanism. Both architectures are taking the joint node 1, the right wrist as an example. Other nodes refer to Fig. 3.

The Channel-squeezing Block: To capture the relevant information specific to PT and filter noise, we hypothesize that (i) the short-range joints (i.e., on the same side of the body) contain slightly less relevant information compared with the target joint itself based on the information gain analysis; (ii) the long-range nodes (i.e., not on the same side of the body) contains much less information relevant to PT classification based on the clinician observation [2,9]. Hence, we propose the following channel-squeezing operation to reach the above hypothesis:

Suppose node m to be the target node, node k to be the relevant node of m, such that the shortest path between two nodes in the graph domain is $k-a$. We propose Eq.2 to determine the output channel size of the relevant node k:

$$C_{out,k} = b \times C_{in}, \quad |k-m| \leq 2 \quad and \quad C_{out,k} = d^{|k-m|} C_{in}, \quad |k-m| > 2 \quad (2)$$

where b, d are the channel-squeezing ratios for short-range and long-range node, respectively. $b, d \in [0, 1]$ and $b \gg d$. $C_{out,k}$ is the output channel size of node k. $|\cdot|$ is the distance between node m and k in the graph domain.

The Channel-fusion Block: To fuse the relevancy information of the target joint m from different ranges, we propose a two-stage fusion process to first fuse long-range features from less-related joints by f_l, then fuse all features by f_a:

$$\mathbf{h_m} = f_a[\mathbf{h_{slef}}, \mathbf{h_{short}}, f_l(\mathbf{h_{long,p}})\mathbf{W^m} \quad (3)$$

where $\mathbf{h_{long,p}}$ is features of long-range related node p, $\mathbf{h_{short}}$ and $\mathbf{h_{slef}}$ are features of short-range related nodes and self-range node, respectively. $\mathbf{W^a}$ is the final weight of node m.

Implementation Details: As shown in Fig. 1, we use two GNN blocks (64, 128 output channel size respectively) with each consisting of an LCN layer, a batch normalization layer, an LeakyReLU layer (0.2 alpha), and a dropout layer (0.2 rates). After two GNN blocks, we apply a PCSF block, a global average pooling layer and a fully connected layer. We use the focal-loss [15] as the loss function for overcoming class imbalance in multiclass classification task. The

optimizer is chosen as Adam, and we train the model with a batch size of 16, a learning rate of 0.01 with 0.1 decay rate, and a maximum epoch of 500 for binary classification; For multiclass classification, the learning rate, weight decay, batch size, and epoch are 0.001, 0.1, 500, 8, and 500, respectively. Empirically, we set the short- and long-range channel-squeezing ratios b, d to 0.9 and 0.125, respectively, returns the most consistently good results.

3 Experiments

Our experiments were run on a PC with Ubuntu 18.04 and an NVIDIA GeForce RTX 3080. Our system is low-cost as it only requires an average GPU memory usage of 1.48 gigabytes for training. The total model training time on the TIM-TREMOR dataset is about ten hours, including human pose features extractions from RGB videos. It only takes about 48s for the PT classification of 1000 frames 30FPS video recording(∼33s), which can be employed in interactive-time diagnosis.

The dataset: We verify our model on a publicly available TIM-TREMOR (Technology in Motion Tremor Dataset) dataset [23]. The dataset consists of 917 video recordings from 55 participants sitting in a chair and performing a set of 21 tasks, and videos range from 18 seconds to 112 seconds. There are 579 videos that present different types of tremors, including 105 PT, 182 Essential Tremor (ET), 88 Functional Tremor (FT), and 204 Dystonic Tremor (DT) videos. Another 60 videos have no tremor during the assessment. The remaining 278 videos with ambiguous diagnosis results are labeled as "Other".

Setup: We first eliminated inconsistent videos to avoid label noise, that is, (i) videos with motion tasks recorded only on a minor subset of participants; (ii) videos with ambiguous diagnosis label -"other". Then, we clip each video into samples of 100 frames each, with the number of clips depending on the length of the successive video frames where the participant is not occluded by the interaction with the clinician. Each clip inherits the label of the source video and is considered an individual sample. A voting system [16,17] is employed to obtain the video-level classification results. This clipping-and-voting mechanism increases the robustness of the system and augments the sample size for training. We employ a 5-fold cross-validation to evaluate our proposed system.

To evaluate the generalizability of the proposed method, we validate our system not only on the binary classification (i.e., classify PT label with non-PT labels), but also on a more challenging multiclass classification task that classifies samples with five tremor labels (PT, ET, FT, DT, and No tremor). We report the mean and standard deviation among all cross-validation for the following metrics: the metrics for the binary classification includes the accuracy (AC), sensitivity (SE), specificity (SP), and F1-Score; the metrics for the multiclass classification are AC and per-class and macro average F1-score, SE and SP.

Table 1. The comparisons on the binary classification (PT v.s. non-PT) task and the summarized multiclass classification (PT v.s. ET v.s DT v.s FT v.s non-tremor) results.

		Binary Classification			
Method		AC	SE	SP	F1
CNN-LSTM [28]		81.0	n/a	79.0	80.0
LSTM [28]		80.0	n/a	79.0	79.0
SVM-1 [28]		53.0	n/a	63.0	55.0
ST-GCN[31]		87.7 ± 3.8	88.3 ± 5.3	87.4 ± 3.1	87.0 ± 4.4
CNN-Conv1D		81.6 ± 5.7	83.4 ± 9.1	80.7 ± 4.4	80.3 ± 6.0
Decision Tree		74.5 ± 4.7	73.4 ± 5.7	75.8 ± 4.0	73.6 ± 4.6
SVM		64.3 ± 5.4	62.2 ± 7.5	66.7 ± 4.6	63.1 ± 7.1
Ours	SPAPNet - full	**90.9 ± 3.4**	**90.7 ± 5.0**	**91.3 ± 2.3**	**90.6 ± 3.7**
	w/o PCSF	88.4 ± 4.5	90.4 ± 6.9	87.0 ± 3.7	87.5 ± 5.2
	w/o Attention	82.6 ± 5.3	82.7 ± 6.0	82.8 ± 5.1	81.3 ± 6.8
		Multiclass Classification			
ST-GCN [31]		70.3 ± 6.9	69.5 ± 6.4	90.7 ± 5.4	67.9 ± 6.7
CNN-Conv1D		63.1 ± 6.5	59.5 ± 5.6	90.8 ± 7.4	61.9 ± 8.3
Decision Tree		54.3 ± 5.7	49.0 ± 7.3	92.3 ± 5.4	55.5 ± 6.5
SVM		47.6 ± 6.4	45.7 ± 6.9	91.6 ± 6.1	52.1 ± 7.2
Ours	SPAPNet - full	**73.3 ± 6.8**	**72.8 ± 5.1**	**92.3 ± 4.1**	**70.7 ± 6.5**
	w/o PCSF	69.1 ± 6.9	69.9 ± 4.0	88.2 ± 4.6	65.7 ± 7.1
	w/o Attention	65.9 ± 6.8	64.2 ± 5.5	90.4 ± 7.9	65.0 ± 7.9

Comparison with Other Methods: To evaluate the effectiveness of our system, we compare our results with the following state-of-the-art video-based PT classification methods: (i) CNN-LSTM [28]: This method uses a CNN-LSTM model to classify the PT and non-PT classes from hand landmarks extracted by MediaPipe [32], their data is videos from the TIM-TREMOR dataset; (ii) SVM-1 [28]: This is a support vector machine model proposed to classify the PT and non-PT classes by the same features in [28]; (iii) LSTM [28]: This is an LSTM deep neural network proposed to classify the PT and non-PT classes by the same features in [28]; (iv) ST-GCN [31]: This is a spatial and temporal graph convolutional neural network for classification tasks on human pose data. For works in [28], we only report the performance in their work since the source code is not publicly available. To compare the effectiveness of our system with conventional methods, we implement a CNN with 1D convolutional layers (CNN-Conv1D) [28] and two machine learning-based methods, namely Decision Tree (DT) and SVM.

From the binary classification result in Table 1, our full system outperforms state-out-of-the-arts [28,31] and other implemented methods. Our AC, SE, SP, and F1 achieves over 90% with standard deviations less than 5%, which indicates the effectiveness and robustness in classifying PT class with non-PT class. Our system achieves better performance by only applying spatial convolution instead of a more deep architecture like spatial-temporal convolution modeling method, ST-GCN [31]. The result validates that our proposed PCSF block effectively improves classification performance and mitigates the overfitting risk in small datasets. Moreover, although our system is designed for binary classification purposes, the full system also shows effectiveness and generalizability by outperforming others in the multiclass classification task. The high macro-

average SP showed relatively reliable performance in identifying people without corresponding tremor labels. Improving the multiclass classification AC and SE is scheduled in our future work.

Ablation Studies: We perform an ablation to evaluate whether there is any adverse effect caused by the proposed PCSF block or the whole attention module. From the rows of "Ours" in Table 1, we observe the effectiveness of the PCSF block and attention module from the performance reduction across all metrics when eliminating the PCSF or the whole attention module for both classification tasks. In addition, we observe the stability of using the full system as it has smaller standard deviations than its variants. Besides, we can observe that the vanilla GNN (i.e., SPAPNet w/o Attention) presents better performance than CNN-Con2D in both classification tasks. It demonstrates the effectiveness of learning human pose features in the graph domain. Moreover, the results show the advantage of deep learning networks by comparing them with two machine learning-based methods, which are decision tree and SVM.

Qualitative Analysis: Fig. 3a. visualizes the interpretability of our system by presenting the mean attention weights of each skeleton joint among all cross-validation. We notice that the mean attention weights of 'Right Wrist' and 'Left Wrist' are significantly higher than others on both classification tasks. It indicates our system pays more attention to the movements of participants' wrists. In addition, the attention weight of 'Neck' is lower than others significantly. One possible reason is that the participants are sitting on the chair, and their neck joint has the smallest global variance during the whole video.

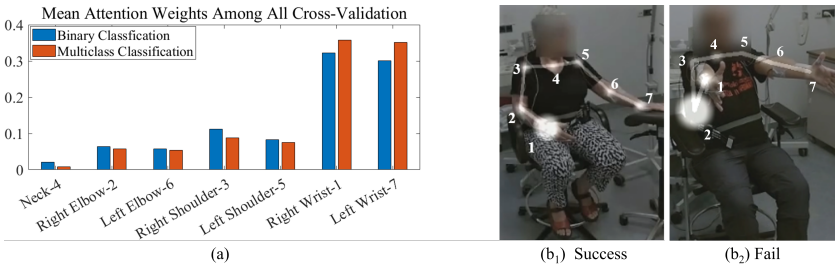

Fig. 3. (a) The mean attention weights of different joints among all cross-validation for both classification tasks; (b) The visualization of the attention weights at a single example frame. The joint index numbers in (b) corresponds to (a); (b_1) One frame in a successful diagnosis; (b_2) One frame in a false diagnosis.

We also analyze the situation in which our method fails or succeeds. Fig. 3 b_1. is a frame in a successful diagnosed example of a PT patient. Consistent with the clinician PT diagnosis based on right hand resting tremor, the right wrist node contributes the most attention. Fig. 3 b_2. is a frame in misdiagnosis,

and the attention is incorrectly dominated by the mis-detected joint position of the right elbow from the pose extraction algorithm. Therefore, it highlights the importance of improving pose extraction performance for future work.

4 Conclusion

In this work, we propose a novel interpretable method SPAPNet to diagnose Parkinson's from the consumer-grade RGB video recordings. Our system outperforms state-of-the-arts by achieving an accuracy of 90.9% and an F1-score of 90.6%. The proposed attention module aids in improving both classification performance and system interpretability. Our proposed novel lightweight pyramidal channel-squeezing-fusion block effectively learns the self, short and long-range relevant information specific to Parkinson's tremor and filters irrelevant noise. Our system shows the potential to support non-intrusive PD diagnosis from human pose videos. Since our system only requires the consumer-grade human pose videos as input, it provides a way for diagnosis of PD in the resource-limited regions where the clinical experts are not abundant. In addition, our system shows potential for remote diagnosis of PD in special situations (e.g., COVID-19 epidemic) and automatic monitoring of PT symptoms during daily life for PD diagnosis.

References

1. Alle, S., Priyakumar, U. D.: Linear Prediction Residual for Efficient Diagnosis of Parkinson's Disease from Gait. In: MICCAI. (2021).
2. Bhat, S., Acharya, U. R., Hagiwara, Y., Dadmehr, N., Adeli, H.: Parkinson's disease: Cause factors, measurable indicators, and early diagnosis. In: Computers in Biology and Medicine. vol. 102, pp. 234-241. (2018).
3. Beitz, J. M.: Parkinson's disease: A review. In: Front Biosci (Schol Ed). 6: pp.65–74. (2014)
4. Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E. and Sheikh, Y.: OpenPose: Realtime multi-person 2d pose estimation using part affinity fields. iN: arXiv e-prints, p. arXiv:1812.08008. (2018).
5. Chen, C., Ramanan, D.: 3D Human Pose Estimation = 2D Pose Estimation + Matching. In: the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7035-7043. (2017)
6. Ci, H., Ma, X., Wang C., Wang, Y.: Locally Connected Network for Monocular 3D Human Pose Estimation. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 3, pp. 1429-1442. (2022)
7. Vásquez-Correa, J. C., Arias-Vergara, T., Orozco-Arroyave, J. R., Eskofier, B., Klucken, J. and Nöth, E.: Multimodal Assessment of Parkinson's Disease: A Deep Learning Approach. In: IEEE Journal of Biomedical and Health Informatics, vol. 23, no. 4, pp. 1618-1630. (2019).
8. Li, S., Gao, Z., Lin, H.: Lookhops: light multi-order convolution and pooling for graph classification. In: arXiv preprint arXiv:2012.15741. (2020).
9. Fahn, S.: Description of Parkinson's Disease as a Clinical Syndrome. In: Annals of the New York Academy of Sciences. vol. 991, pp. 1-14. (2003).

10. Gibb, W. R., Lees, A. J.: The relevance of the Lewy body to the pathogenesis of idiopathic Parkinson's disease. In: J Neurol Neurosurg Psychiatry.51:745–52.(1988).
11. Hausdorff J. M.: Gait dynamics in Parkinson's disease: common and distinct behavior among stride length, gait variability, and fractal-like scaling. In: Chaos (Woodbury, N.Y.), 19(2), 026113. (2009).
12. Hssayeni, M. D., Jimenez-Shahed, J., Burack, M. A., Ghoraani, B.: Wearable Sensors for Estimation of Parkinsonian Tremor Severity during Free Body Movements. In: Sensors (Basel, Switzerland), 19(19), 4215. (2019).
13. Kipf, N. and Welling, M.: Semi-supervised classification with graph convolutional networks. In: ICLR. (2017).
14. Patel, S., Lorincz, K., Hughes, R. *et al.*: Monitoring Motor Fluctuations in Patients With Parkinson's Disease Using Wearable Sensors. In: IEEE Transactions on Information Technology in Biomedicine, vol. 13, no. 6, pp. 864-873, Nov. (2009).
15. Lin, T. Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: CVPR. pp. 2980–2988. (2017).
16. Lu, M., Poston, K., Pfefferbaum, A., Sullivan, E. V., Fei-Fei L, Pohl, K. M., Niebles, J. C., Adeli, E.: Vision-based Estimation of MDS-UPDRS Gait Scores for Assessing Parkinson's Disease Motor Severity. In: Med Image Comput Comput Assist Interv (MICCAI). (2020).
17. Lu, M., Zhao, Q., Poston, K. L., Sullivan, *et al.*: Quantifying Parkinson's disease motor severity under uncertainty using MDS-UPDRS videos. In: Medical Image Analysis, vol. 73. (2021).
18. Luvizon, D. C., Picard, D., Tabia, H.: 2D/3D Pose Estimation and Action Recognition Using Multitask Deep Learning. In: the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5137-5146. (2018).
19. Massano, J., Bhatia, K.P.: Clinical approach to parkinson's disease: features, diagnosis, and principles of management. In: Cold Spring Harbor Perspectives Med. 2(6), a008870. (2012).
20. Mhyre, T. R., Boyd, J. T., Hamill, R. W., Maguire-Zeiss, K. A.: Parkinson's disease. In: Sub-cellular biochemistry, 65, 389–455. (2012).
21. Mostafa, S. A., Mustapha, A., Mohammed, M. A., Hamed, R. I., Arunkumar, N., Ghani, M., Jaber, M. M., Khaleefah, S. H.: Examining multiple feature evaluation and classification methods for improving the diagnosis of Parkinson's disease, In: Cognitive Systems Research, vol. 54, pp.90-99. (2019).
22. Pasquini, J., Ceravolo, R., Qamhawi, Z., Lee, J., Deuschl, G., Brooks, D. J., Bonuccelli, U., Pavese, N.: Progression of tremor in early stages of Parkinson's disease: a clinical and neuroimaging study. In: Brain. vol 141, issue 3, pp 811–821. (2018).
23. Pintea, S. L., Zheng, J., Li, X., Bank, P., van Hilten, J. J., van Gemert, J. C.: Hand-tremor frequency estimation in videos. In: ECCV Workshops (6), vol. 11134, pp. 213–228. (2018).
24. Rizek, P., Kumar, N., Jog, M. S.: An update on the diagnosis and treatment of Parkinson disease. In: CMAJ : Canadian Medical Association journal, 188(16), 1157–1165. (2016).
25. Rizzo, G., Copetti, M., Arcuti, S., Martino, D., Fontana, A., Logroscino, G.: Accuracy of clinical diagnosis of Parkinson disease: A systematic review and meta-analysis. In: Neurology. 9; 86(6): 566-76. (2016)
26. Sveinbjornsdottir, S.: The clinical symptoms of Parkinson's disease. In: J. Neurochem., 139: 318-324. (2016).
27. Wang, J., Yan, S., Xiong, Y., Lin, D.: Motion Guided 3D Pose Estimation from Videos. In: Computer Vision – ECCV 2020. Lecture Notes in Computer Science, vol 12358. Springer, Cham. (2020).

28. Wang, X., Garg, S., Tran, S.N. *et al.*: Hand tremor detection in videos with cluttered background using neural network based approaches. In: Health Inf Sci Syst. 9, 30 (2021).
29. Wang, W., Lee, J., F. Harrou, F., Sun, Y.: Early Detection of Parkinson's Disease Using Deep Learning and Machine Learning. In: IEEE Access, vol. 8, pp. 147635-147646, (2020).
30. Wirdefeldt, K., Adami, H. O., Cole, P., Trichopoulos, D., Mandel, J.: Epidemiology and etiology of Parkinson's disease: a review of the evidence. In: Eur J Epidemiol. 26 Suppl 1:S1-58.Jun. (2011).
31. Yan, S., Xiong, Y., Lin, D.: Spatial temporal graph convolutional networks for skeleton-based action recognition. In: AAAI Conference on Artificial Intelligence. (2018).
32. Zhang, F., Bazarevsky, V., Vakunov, A., Tkachenka, A., Sung, G., Chang, C. L., Grundmann, M.: MediaPipe Hands: On-device Real-time Hand Tracking. In: arXiv preprint. arXiv:2006.10214. (2020).
33. Zhang, L., Wang, M., Liu, M., Zhang, D.: A Survey on Deep Learning for Neuroimaging-Based Brain Disorder Analysis. In: Frontiers in Neuroscience. vol. 14. (2020).