An Automatic Evaluation Method for Parkinson's Dyskinesia Using Finger Tapping Video for Small Samples

Zhu Li
Hangzhou Dianzi University https://orcid.org/0000-0001-6373-5846
lu kang
Hangzhou Dianzi University
Miao Cai (✉ caimiao1985@126.com)
Zhejiang Hospital
Xiaoli Liu
Zhejiang Hospital
Yanwen Wang
Zhejiang Hospital
Jiayu Yang
Hangzhou Dianzi University

Research Article

Keywords: Parkinson's disease, FCN, Pose estimation, Manual feature, small sample classification

Posted Date: January 11th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1207003/v1

An automatic evaluation method for Parkinson's dyskinesia using finger tapping video for small samples

Zhu Li[1] • Kang Lu[1] • Miao Cai[2] • Xiaoli Liu [2] • Yanwen Wang[2] • Jiayu Yang[1]

[1] School of Electronics and Information Engineering, Hangzhou Dianzi University, Hangzhou, Zhejiang, China.

[2] Neurology department, Zhejiang hospital, Hangzhou, Zhejiang, China.

Correspondence should be addressed to Miao Cai; caimiao1985@126.com

Abstract

Purpose The assessment of dyskinesia in Parkinson's disease (PD) based on Artificial Intelligence technology is a significant and challenging task. At present, doctors usually use MDS-UPDRS scale to assess the severity of patients. This method is time-consuming and laborious, and there are subjective differences. The evaluation method based on sensor equipment is also widely used, but this method is expensive and needs professional guidance, which is not suitable for remote evaluation and patient self-examination. In addition, it is difficult to collect patient data in medical research, so it is of great significance to find an objective and automatic assessment method for Parkinson's dyskinesia based on small samples.

Methods In this study, we design an automatic evaluation method combining manual features and convolutional neural network (CNN), which is suitable for small sample classification. Based on the finger tapping video of Parkinson's patients, we use the pose estimation model to obtain the action skeleton information and calculate the feature data. We then use the 5-folds cross validation training model to achieve optimum trade-of between bias and variance, and finally make multi-class prediction through fully connected network (FCN).

Results Our proposed method achieves the current optimal accuracy of 79.7% in this research. We have compared with the latest methods of related research, and our method is superior to them in terms of accuracy, number of parameters and FLOPs.

Conclusion The method in this paper does not require patients to wear sensor devices, and has obvious advantages in remote clinical evaluation. At the same time, the method of using motion feature data to train CNN model obtains the optimal accuracy, effectively solves the problem of difficult data acquisition in medicine, and provides a new idea for small sample classification.

Keywords Parkinson's disease · FCN · Pose estimation · Manual feature · small sample classification

1 Introduction

Parkinson's disease is the second most common neurodegenerative disease followed by Alzheimer's disease [1]. The patients are mainly manifested by static tremor, muscle stiffness, bradykinesia and postural instability [2,3]. Accurate and objective evaluation results of Parkinson's disease should be obtained in the treatment of Parkinson's disease. At present, there are many ways to evaluate the motor function of patients with PD, among which the MDS-UPDRS scale[4], as a standard rating scale for PD evaluation, is widely adopted in the evaluation of PD motor level, because of its simplicity and comprehensiveness, it is widely used in evaluating the motor level of patients with PD[5]. However, the accuracy of scale-based evaluation directly depends on doctors' clinical experience, and has subjective differences, so it is of great significance to provide an objective and automatic PD evaluation method to assist doctors' clinical evaluation and the self-examination of patients.

Thanks to clinical needs and the rapid development of deep learning, CNN is widely used in the evaluation and diagnosis of PD. The research is focused on the assessment of motor disorders, pathological analysis and early diagnosis of PD [6-8]. Dyskinesia is the core symptom of Parkinson's disease, so this paper focuses on the related research of motor dysfunction in PD. At present, the primary research method based on deep learning is to obtain characteristic data through sensor devices for monitoring, analysis and evaluation [9-11]. For example, the system based on body network sensor proposed by Parisi F et al. [12]can automatically evaluate the severity of PD patients' gait by extracting kinematic features in time domain and frequency domain to characterize the Parkinson's disease gait of

PD patients' gait. The detection system based on mechanical impedance proposed by Dai H et al. [13] quantitatively evaluates myotonia in patients with PD by extracting body movement information. The research method based on sensors or wearables, which can obtain more deep and complex information [14], has great potential in the quantitative evaluation of PD and the development of treatment equipment, however, it depends on the guidance of professionals, the equipment restricts the movement of patients to a certain extent, with limited application scenario, so it is not applicable to the regular evaluation of PD patients.

It is a potential method to use the model of action recognition to automatically evaluate PD without professional equipment. At present, the CNN-based action recognition has demonstrated good performance, especially in skeleton dataset [15,16], many action recognition tasks have achieved perfect results. For example, the method proposed by Li C et al. [17]has achieved the accuracy of 92.08% on the skeleton-based ChaLearn gesture dataset. Y et al. [18]proposed a channel-wise topology refinement graph convolution network, which has achieved outstanding action recognition performance on NTU RGB + D120 dataset, the accuracy of cross-subject and cross-view is 88.9% and 90.6% respectively. Despite good performance demonstrate, the above methods are focused on coarse-grained tasks, most of which are used to identify activity scenes with prominent action differences, and have difficulty in dealing with complex tasks [19,20]. In addition, the model architecture of action recognition is complex, and the model training depends on a large number of datasets, but it is difficult to obtain datasets in medical research, especially the clinical data based on patients' behavior. Because of the limitation of the number of patients and their privacy protection, the data volume collected often cannot help in the training of complex models, thus it is necessary to find a method based on small samples to achieve the evaluation of PD.

Finger tapping test is often used to evaluate the motor dysfunction in PD [21]and neurophysiological examinations [22], since the motor characteristics of finger tapping are closely correlated with bradykinesia [23,24]. The motor function examination in the MDS-UPDRS includes the finger tapping test. And the MDS-UPDRS provides the grading standard of finger tapping test as a reference for doctors to evaluate PD, which also provides a theoretical basis for this paper. According to

Goetz CG et al. [25], the accurate scoring and interpretation of the results of finger tapping test requires a wealth of experience, so it becomes one of the most difficult items to evaluate in PD motor examination, and even for professional doctors, subtle differences in motors are difficult to detect. By observing the evaluation criteria of finger tapping, it can be found that the difference of finger tapping test between adjacent scores (for example, scored 1 and 2) is fuzzy. This high-fine-grained recognition task poses a severe challenge to the classification model based on CNN. This research is conducted based on the high fine-grained task of finger tapping test.

In view of the above problems, based on Parkinson's video data set, this paper designs a method of combining manual features with CNN to realize the evaluation and classification of Parkinson's finger tapping test. This paper mainly includes the following aspects:

i) Sensors and other auxiliary equipment are inconvenient to wear, need special guidance, and cannot record the disease changes of PD patients continuously and promptly. The evaluation method in this paper extracts the feature data by analyzing the video dataset automatically, and then evaluates and diagnoses it with CNN. The dataset can be photographed by smart phone and does not depend on sensor.

ii) In view of small differences in finger tapping test and difficult to distinguish, this paper summarizes the change law of finger tapping test, designs a feature based on range and velocity of the action. Firstly, this paper uses the pose estimation algorithm to extract the hand skeleton data, then calculates the feature data based on the hand skeleton data, and finally evaluates and grades the action through CNN, which provides a new idea for PD evaluation.

iii) The combination of manual feature and CNN designed in this paper effectively solves the problem of difficult data acquisition in medical research. Compared with the dependence of previous action recognition models on a large number of data sets, this method can also achieve good performance in the case of small samples.

The other parts of this paper are arranged as follows: In the second part, the method of combining manual features and CNN designed in this paper is introduced in detail; In the third part, the experiments are presented, including dataset preparation, experimental results of the methods, comparison experiment with the existing methods and analysis experiment for manual features; In the fourth part, the methods and experimental results are discussed and analyzed; Finally, the fifth part is a conclusion of this paper.

2 Methods

The flow chart of the finger tapping evaluation method proposed in this paper is shown in figure 1. Firstly, the finger tapping test video of PD patients is collected, and then the skeleton data of the hands is extracted by the pose estimation model Mediapipe Hands [26]. Subsequently, the feature data based on the motion law of the hands is extracted depending on the method designed in this paper, so as to obtain the one-dimensional time series data. Following the data pre-processing such as normalization and cropping alignment, the data are input into the FCN, and finally the output of five-classification results is obtained, which is the score prediction of the corresponding finger tapping.

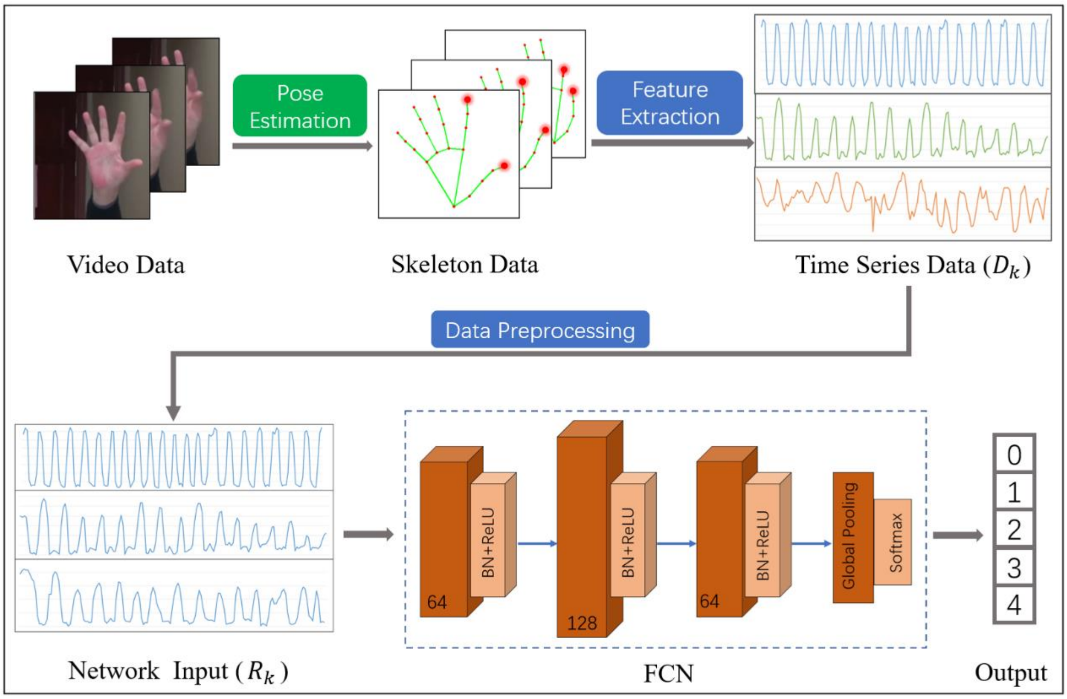


Fig. 1 The flow chart of the methods in this paper

2.1 Pose estimation

In this paper, the Mediapipe algorithm is used to extract the hand skeleton data from the video with finger tapping of PD patients. Mediapipe Hands is one of the most advanced frameworks for hand skeleton estimation, which is robust to partially visible and occluded hand estimation. It detects the skeleton of the hands by the two models working together: 1) Palm Detection Model, which searches the whole image and returns the predicted hand boundary box; 2) Hand Landmark Model, which detects the area of the hands returned by the palm detector and returns the high-fidelity key points of the hands. The key hand points returned by the model contain the 3D coordinates of 21 knuckles, and the position and name of the corresponding knuckles are shown in figure 2.

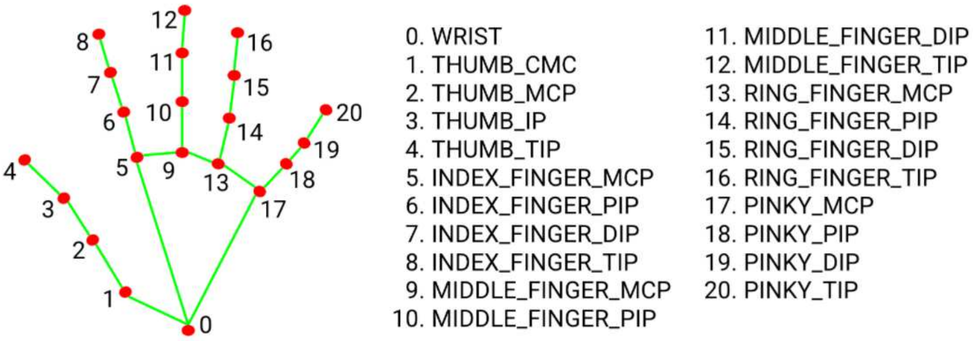


Fig. 2 The location and name of the 21 3D coordinates returned by MediaPipe Hands

We use the Mediapipe Hands method to detect the video data of PD patients, and the effect is shown in figure 3, from which we get the hand skeleton data of the finger tapping of all patients. Each video frame corresponds to a set of three-dimensional data of joints. We use the 3D coordinate $J_i = \{x_i, y_i, z_i\}$ to represent the joint i. Suppose there are T-frame pictures in each video, and each hand skeleton includes u joint points. In this paper, u=21 means the hand feature M_t of frame t can be expressed as $M_t = \{J_1^t, J_2^t, \cdots, J_u^t\}$.

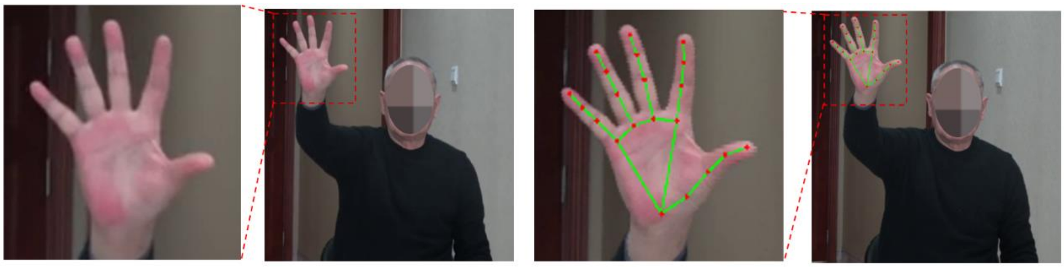


Fig. 3 The effect of extracting hand key points in patients' finger tapping by Mediapipe Hands.

2.2 Manual feature extraction

In this paper, the feature extraction method is designed based on the action evaluation standard of MDS-UPDRS. The analysis of the evaluation index of finger tapping, we can find that the difference of different scores is mainly reflected in the opening range, tapping velocity and interruption behavior of fingers, so we designed the following feature extraction methods to characterize their action rules.

Section 2.1 presents that the three-dimensional data of the hand skeleton is obtained using Mediapipe. In order to calculate the range and velocity changes of finger tapping, this paper selects the three-dimensional data of joint 4.THUMB_TIP and joint 8.INDEX_FINGER_TIP, which are $J_4 = \{x_4, y_4, z_4\}$ and $J_8 = \{x_8, y_8, z_8\}$ respectively. By calculating the Euclidean distance between the two joints, the fingertip distance data of frame t is obtained, which are shown as follows:

$$D_{\mathrm{t}} = distance(J_4^t, J_8^t) = \sqrt{(x_4^t - x_8^t)^2 + (y_4^t - y_8^t)^2 + (z_4^t - z_8^t)^2} \tag{1}$$

The D_t is extracted based on the pixel distance of the image. Since the change of shooting distance and camera jitter may cause errors, this paper uses Z-score standardization processing to get the standardized data. Firstly, the μ and σ of D_{t} are calculated.

$$\mu = \frac{D_1 + D_2 + \cdots + D_T}{T} = \frac{\sum_{t=1}^{T} D_t}{T} \quad , \quad \sigma = \sqrt{\frac{\sum_{\mathrm{t}=1}^{T} (D_t - \mu)^2}{T}} \tag{2}$$

Then the standardized data of frame t can be expressed as:

$$S_t = \frac{D_{\mathrm{t}} - \overline{D}}{\sigma} \tag{3}$$

The one-dimensional sequence data of the k-th video can be expressed as $R_k = \{S_1^k, S_2^k, \cdots, S_T^k\}$, that is, the final model input data shown in this paper, and time series classification has always been a classical task in deep learning. The feature data extracted in this paper accords with the law of time series data, with a good potential by using CNN processing.

In order to test the variation law of the range and velocity of finger tapping separately, the range and velocity data of ten finger tappings are extracted to verify the experimental results. The range and velocity are expressed as A_k and B_k, respectively as follows:

$$A = \{a_1, a_2, \cdots, a_{10}\} = \{(\mathrm{m}_1 - \mathrm{n}_1), (\mathrm{m}_2 - \mathrm{n}_2), \cdots, (\mathrm{m}_{10} - \mathrm{n}_{10})\} \tag{4}$$

$$B = \{b_1, b_2, \cdots, b_{10}\} = \{(t_2^m - t_1^{\mathrm{m}}), (t_3^m - t_2^m), \cdots, (t_{11}^m - t_{10}^m)\} \tag{5}$$

We use a_i and b_i to represent the range and velocity of the i-th action, $\mathrm{i} = 1, 2, \cdots, 10$, as shown in figure 4, m_i to m_{i+1} is a complete action cycle. m_i and n_i represent the maximum and minimum of the i-th action, respectively. And $\mathrm{t}_{i+1}^m - t_i^m$ represent the velocity of the i-th action.

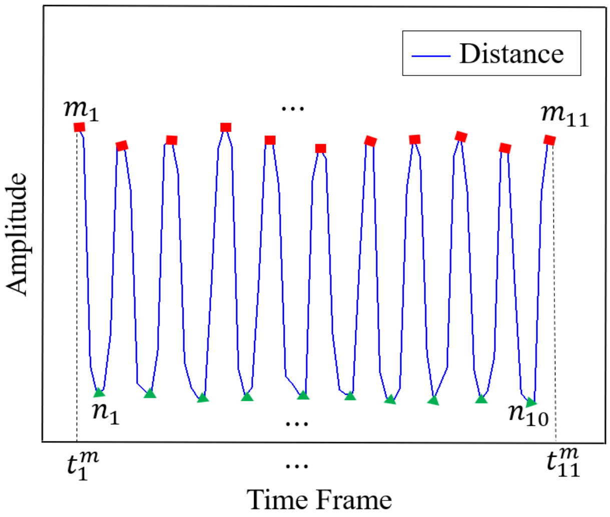


Fig. 4 A schematic diagram for the data of finger tapping

2.3 Model framework

The modified FCN model framework is shown in figure 5. As a sequence classification model, the input of the model is one-dimensional sequence data, and the overall structure is composed of three convolutional layers and a global average pooling (GAP) layer. The convolution layer is used for feature extraction, each layer of convolution output connects a batch normalization and the ReLU activation function. And the GAP layer is used for classification, followed by the Softmax activation function. In this paper, the model uses the global average pooling layer instead of the full connection layer, which can accept the sequence of any dimension, better correspond the category to the feature map of the last convolutional layer, so as to achieve accurate classification results. The model is modified according to the characteristics of the self-made dataset. The details of the modified model are shown in Table 1.

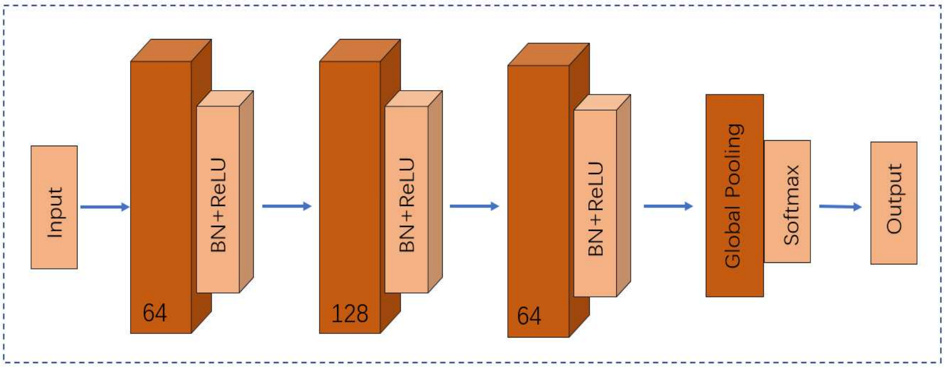


Fig.5 The architecture of the modified FCN model, where 64 and 128 is the number of channels of feature map

Table 1 Layers information for proposed FCN architecture

No.	Layer	Information	
1.	Input layer	Size	180×1
2.	Con_1	Number of filters	64
		Kernel size	8×1
3.	Batch_Norm_1	Number of channels	64
		Activation	ReLU
4.	Con_2	Number of filters	128
		Kernel size	5×1
5.	Batch_Norm_2	Number of channels	128
		Activation	ReLU
6.	Con_3	Number of filters	64
		Kernel size	3×1
7.	Batch_Norm_3	Number of channels	64
		Activation	ReLU
8.	GAP	Size	5
		Activation	Softmax

3 Experiment and Results

This section mainly presents the experimental process and experimental results, and the experimental preparation introduces the datasets used in this paper and the main indicators for measuring the performance of the CNN. The experiment consists of four parts. The first part is the behavior recognition experiment. In this paper, the classification of finger tapping is regarded as a action recognition task, and the shortcomings of action recognition method are analyzed; The second part introduces the experiment based on manual features designed in this paper, which mainly presents the experimental details and results of this method; The third part illustrates the performance and advantages of this method depending on lots of comparative experiments; Finally, the fourth part presents the experiment based on the manual feature designed in this paper to verify the rationality and reliability of our method.

3.1 Experimental preparation

The dataset was collected from Zhejiang Hospital in China, which contained 252 data of 120 people. The dataset is the video of finger tapping of each person's both hands. All the included data have obtained the informed written consent of the patients. The sex and age are shown in Table 2.

Table 2 Sex and age distribution of data set

Age group	<50	50-60	60-70	70-80	80+	Total
Male	1	13	28	20	6	68
Female	0	15	21	16	0	52
Total	1	28	49	36	6	120

All the video data used in this research were recorded on ordinary smart phones, and the video frame rate was 30 frames per second (FPS). We clipped the video with the finger tapping of both hands respectively and got a total of 252 videos, with each containing 10 or more finger tappings. There are five levels of truth labels for finger tapping: 0 = normal, 1 = slight, 2 = mild, 3 = moderate, 4 = severe. The evaluation criteria are shown in figure 1. The difference in adjacent scores was difficult to judge, so the true scores of patients were evaluated by Parkinson's disease subspecialists with rich clinical experience. The score distribution is shown in Table 3. Due to the small number of critically ill patients in hospital, there were only two cases classified as 4 in this paper. In the follow-up, we will continue to provide additional patient data for the experiment.

Table 3 Score distribution of data set

Score	0	1	2	3	4	Total
Quantity	54	115	52	30	2	252
Proportion %	21.4	45.7	20.6	11.9	0.4	100

The hardware environment used to test the proposed algorithm is a ubtuntu18.04 system computer with an NVIDIA GeForce GTX 1080Ti (11 GB). Our preprocessing method uses opencv4.4.0 to process data, and the model framework is built on tensorflow1.11.0 platform.

In this paper, four indexes, Accuracy, Precision, Recall and F1-Score, are used to measure the effect of classification and recognition. Accuracy indicates the proportion of the correct results predicted by the model to the total observed values; Precision represents the correct proportion of the results in which the model prediction is a positive example; Recall is the proportion of the correct prediction results in the samples in which the real situation is a positive example; F1-Score is the harmonic average of Precision and Recall, and its values range from 0 to 1. The higher the value, the more accurate the output of the model. The calculation methods of the four indicators are as follows:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{6}$$

$$\text{Pricision}=\frac{TP}{TP+FP} \tag{7}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{8}$$

$$\text{F1-Score}=\frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision+Recall}} \tag{9}$$

Where TP represents the number of samples in which the positive samples are correctly classified; FP means the number of samples in which negative examples are mistakenly identified as positive ones; FN means the number of samples in which positive examples are mistakenly identified as positive ones; The TN means the number of samples that the negative examples are correctly classified.

3.2 Action recognition-based Experiments

Considering that the finger tapping test of patients with different severity of Parkinson's disease have different manifestations, this paper regards the classification of finger tapping test as an action recognition task, and evaluates and classifies PD patients by analyzing action differences. To test the effect of the action recognition method, this paper uses the mature action recognition classification model such as two-stream model and 3DCNN. The 3DCNN model can input video or image frames directly without preprocessing. The two-stream method needs to input image frames and optical flow frames respectively. This paper extracts all video image frames and two-dimensional optical flow frames, as shown in figure 6.

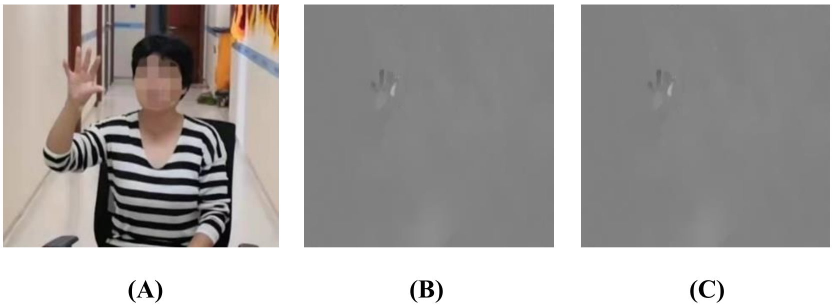
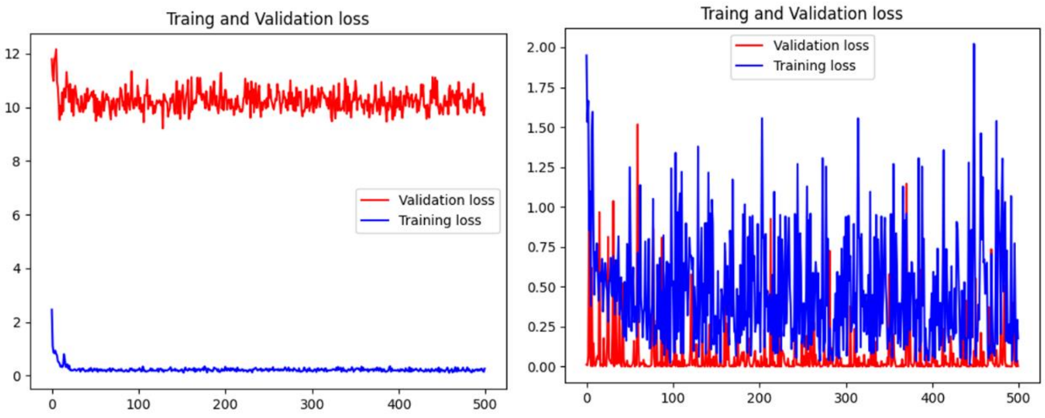


Fig. 6 The image frames and the two-dimensional optical flow frames; **A** for the original image, **B** for the optical flow graph in x direction, and **C** for the optical flow graph in y direction

The experimental results are shown in figure 7, which are Two-stream Fusion [27]and

R2+1D+BERT [28]In the training process of the model, it can be observed that the training processes of the two models are not convergent, and the method based on action recognition can not solve the fine-grained task of finger tapping test. The non-convergence of training is primarily caused by little difference between different scores of finger tapping. The high fine-grained task generally requires a large number of training datasets, the collection of dataset has been a difficult task in medical field, which is a common problem. It is difficult to collect enough data to support the training of complex models in medical research. In addition, the structure of the action recognition model is usually complex, and the detailed features in actions is easily lost in the training process, and this paper needs to extract these detailed features to get better results.

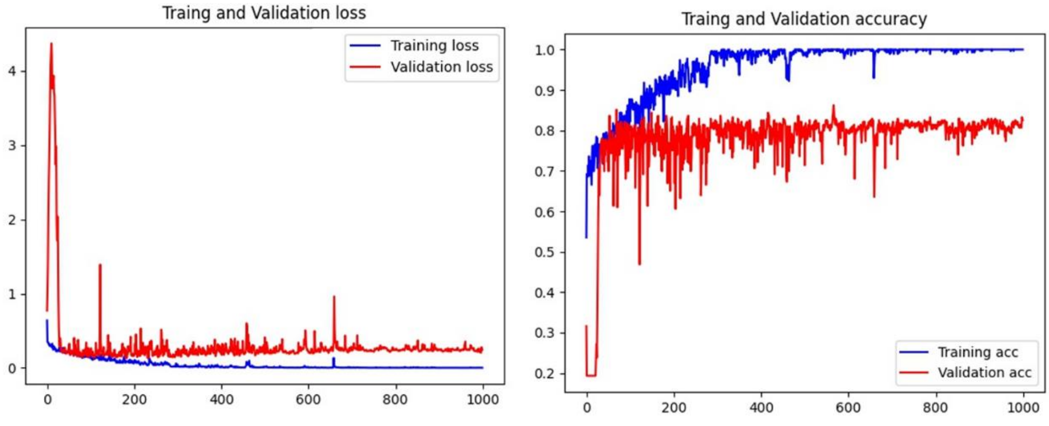


(A) **(B)**

Fig. 7 The loss diagram of action recognition model during the training process; **A** for the Two-stream Fusion model, and **B** for R2+1D+BERT model

3.3 Experiments based on manual features

In order to make up for the difficulty of dataset collection in medicine, Starting from the idea of combining manual design action features with CNN, this paper designs a feature based on the law of finger tapping test. According to the method shown in section 3.2, we extracted the manual feature data of 252 PD patients for training. In this paper, the batch size of the training process is 16, the number of iterations is 1000, and the learning rate is set to le-7, training process as shown in figure 8. We can see that the model training in this paper converges successfully.



(A) **(B)**

Fig. 8 In the training process of the model; **A** for loss, and **B** for the accuracy

Due to the small number of data sets in this paper, in order to obtain reliable and stable model accuracy, this paper uses 5-fold cross validation, which is often used for model training of small-scale data sets, which can optimize the evaluation and selection process of the model. The 5-fold cross validation method is shown in figure 9, which divides the data set into five equal parts, uses the first fold as the test set and the other folds as the training set to get the precision, and in turn, uses the second fold as the test set and other folds as the training set, to get a total of five precision in five times, and averages them to get the model accuracy. As shown in Table 4, we get a classification accuracy of 79.7%, which is the Optimal accuracy in the research based on finger tapping test.

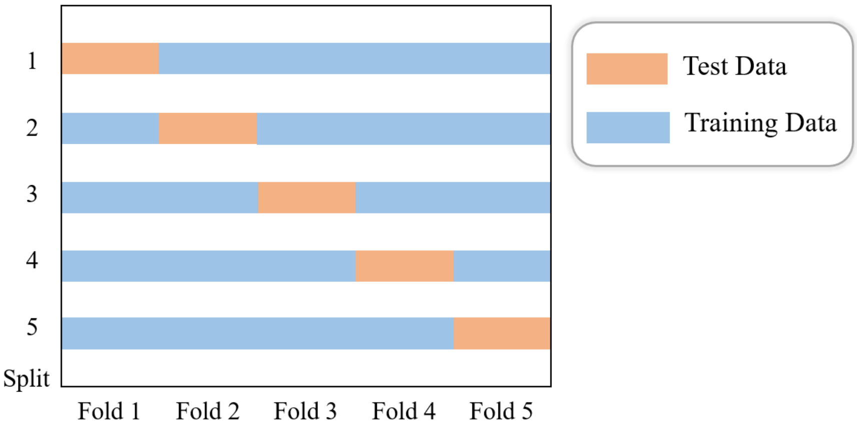


Fig. 9 Schematic diagram of 5-fold cross validation

Table 4 Results of five-fold cross validation results

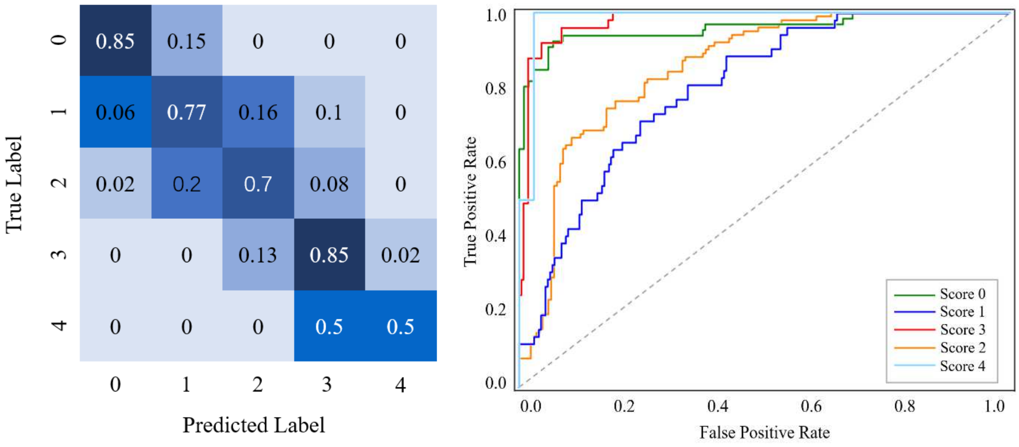
5-fold	1	2	3	4	5	Average
Accuracy %	80.5	77.8	74.1	81.4	84.7	79.7

Table 5 shows the results of four indicators, which are also obtained by the 5-fold cross validated.

The overall results of the indicators reflect the excellent performance of this model. Figure 10 is the confusion matrix based on the classification results, the error recognition rate of each score in the chart is acceptable within the range, and the proportion of prediction errors is mainly reflected in the adjacent scores, showing that the model in this paper has good robustness.

Table 5 Index results of different scores

Score	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
0	79.7	92.6	93.2	92.2
1		78.2	86.4	81.2
2		66.8	49.8	53.8
3		85.4	88.0	86.6
4		50.0	1.00	0.67



(A) **(B)**

Fig. 10 The results of the model; **A** is confusion matrix., and **B** is ROC diagram

3.4 Comparison experiment

The experimental method in this paper is based on manual feature data. As a new method of PD evaluation, there are fewer experimental results to compare with. Table 6 shows that the accuracy of this method is significantly higher than that of other methods by comparing with the existing optimal model based on sensor and skeleton data. In addition, the model used in this paper, as a model for time series classification, has advantages in the number of parameters and FLOPS compared with other models.

Based on the dataset, we also compared different time series models in this paper [32]. As shown in Table 7, the model obtains the best results. Compared with the Resnet model with similar performance, the model in this paper has fewer parameters and obvious advantages in terms of computing resources.

Table 6 Performance comparison with different models

Data type	Model	Acc(%)	Para(M)	Flops(G)
Skeleton data	Motif-GCNs [29]	57.1	1.7	2.2
	2s-AGCN [30]	61.9	6.9	7.9
	Three-Stream [31]	72.4	14.0	3.8
Time series	FCN(ours)	79.7	1.1	1.9

Table 7 Performance comparison with time series models

Model	Acc(%)	Paras(M)	Flops(G)
CNN	66.1	0.9	0.5
Resnet	77.4	2.8	4.8
FCN(ours)	79.7	1.1	1.9

3.5 Feature analysis experiment

Previous experiments have verified the reliability of this method, which shows that our design of manual feature is reasonable and effective, and the range and velocity of finger tapping provide key information for PD evaluation. In order to further verify the role of range and velocity information in the evaluation of PD, we carried out a lot of experiments. First of all, based on the manual feature data of finger tapping, we carried out an unsupervised clustering experiment [33]. The results are shown in figure 11, since the data of patients with a score of 4 in this data set is too small, in order to reduce the interference of the experiment, we excluded the data with a score of 4 and carried out four kinds of clustering experiments.

Although the classification accuracy of unsupervised experiments is not high enough, we can observe that there is a close correlation between the distribution of the four types of data in figure 11 and the evaluation criteria in the MDS-UPDRS, and the action rules of the four types of data are basically consistent with those of the four scores. As shown in figure 11, the range and velocity of finger tapping of (A) are more uniform, and the ranges of (B), (C) and (D) increasingly fluctuate. On this basis, the paper extracts the range and velocity information from the feature data for experiments, that is, the A_k and $B_k$obtained in Section 2.2. The classification results obtained by inputting A_k and B_k into the FCN model are shown in Table 8. It can be found that the accuracy of range data is much higher than that of velocity data, and the experimental results of range data are in line with our expectations, but the classification performance of velocity experiment is poor.

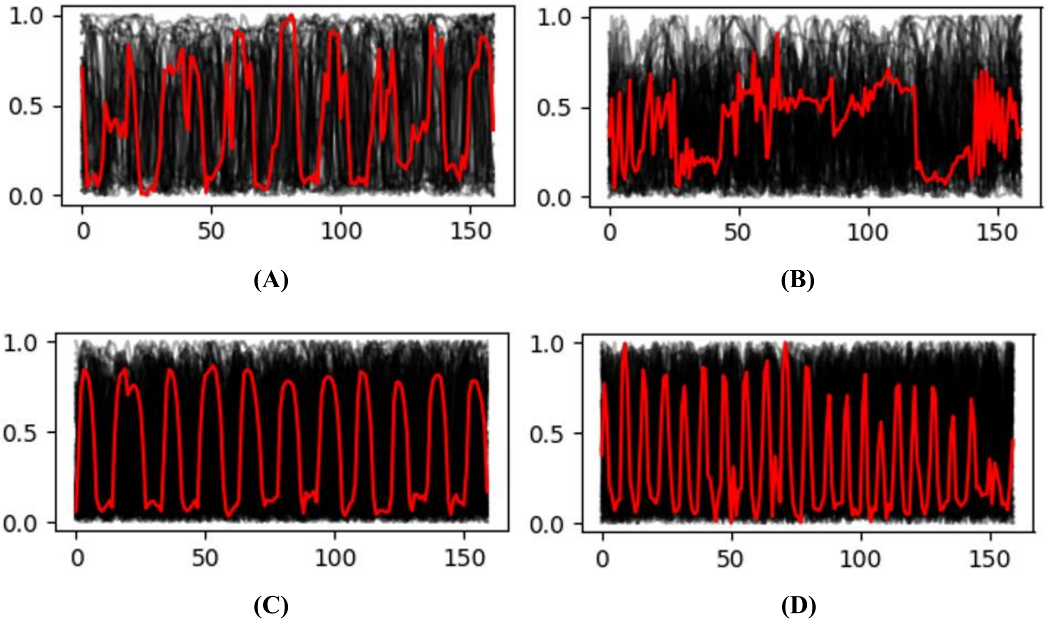


Fig. 11 Clustering chart based on KMedoids+DTW

Table 8 Comparison of range and velocity experiments

	Acc(%)	Dimensions	Data volume
velocity	46.8	10	250
range	68.4	10	250

4 Discussion

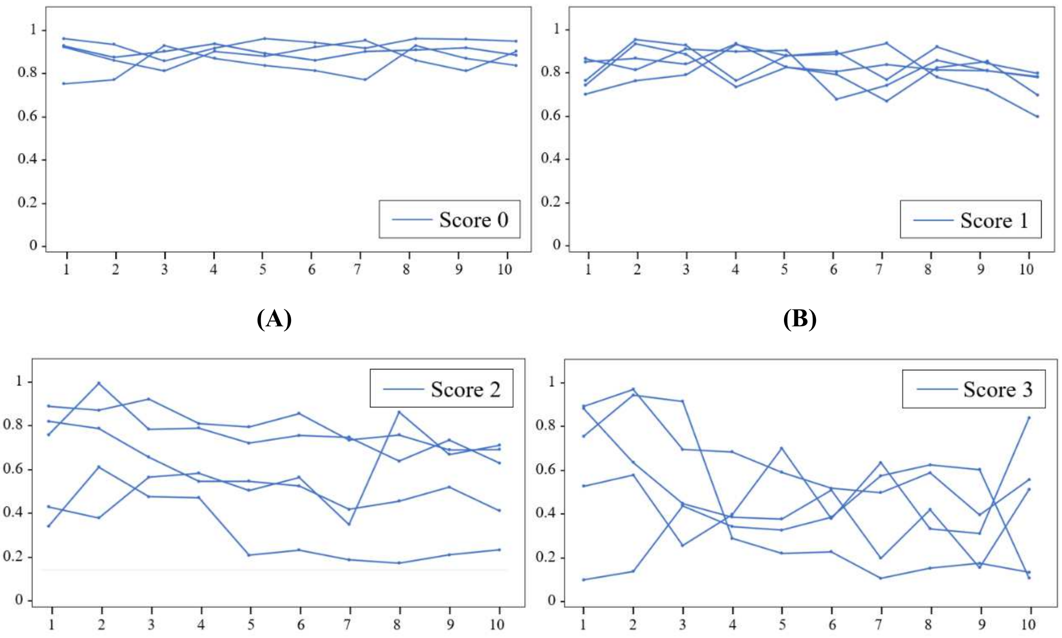
In this paper, we extract self-designed manual features to evaluate the finger tapping test of patients with PD, and preliminarily prove the feasibility of finger tapping test on PD evaluation. The conclusions are obtained based on the experimental results of our self-built dataset. This paper open-sources the code used in the experiment for other researchers to study and verify the results.

First of all, this paper carries out the action recognition experiment based on the original video data. At present, the model training of the mature action recognition model cannot converge, since it is difficult to collect the data set in medical research. The training process of action recognition model requires a large number of datasets, and this method can not solve the high fine-grained task of finger tapping without additional detailed features, who also proves the necessity of extracting skeleton information and designing manual features.

Based on the action evaluation standard of finger tapping in MDS-UPDRS, this paper designs a manual feature, and carry out the training by feature data, which requires a small amount of data, and can effectively solve the problem that the method of action recognition depends on the volume of data. In order to further verify the rationality of the feature design, we first carried out the unsupervised clustering experiment based on the feature data, and found that the clustering distribution of the feature data is basically consistent with that in the MDS-UPDRS, showing that the feature we designed can reflect the motor dysfunction of PD patients, and the method adopted in this paper is reasonable.

To verify the key information in the clustering experiment, we extracted the range and velocity feature data for the experiment, the range feature experiment is basically in line with our expectation, but the velocity feature experiment performance is poor. In order to explain this result, we drew a line chart with different fractional ranges and velocities for analysis by random sampling. We sampled 5 data for each fractional segment in order to facilitate observation, and the specific results are shown in figures 12 and 13. As shown in figure 13, it is found that the range data has some rules, the range value fluctuation gradually becomes larger with the increase of the score, showing a downward trend, which is basically in line with the explanation of the scale. However, no obvious rule is found in the velocity data. We analyzed the methods for collecting the velocity data, and found that the main cause of the error is that we expected to collect the data without professional equipment. However, the FPS of smart phones is low, and the velocity information of captured videos is lost too much, which leads to unsatisfactory results. We will continue to improve in the follow-up experiments to analyze the velocity changes in finger tapping.

To sum up, the method designed in this paper has achieved good results. We have set up a lot of experiments to verify the rationality, reliability and accuracy of the method, and fully discussed and analyzed the experimental results. However, the experiments are also challenged with the problems such as lack of datasets and data imbalance, for example, there are only two cases of data with a score of 4 in the finger tapping experiment, and we can only use four types of data in many experiments, which undoubtedly leads to the incompleteness of our experiment. We will subsequently continue to collect additional data to make the experiment more complete.



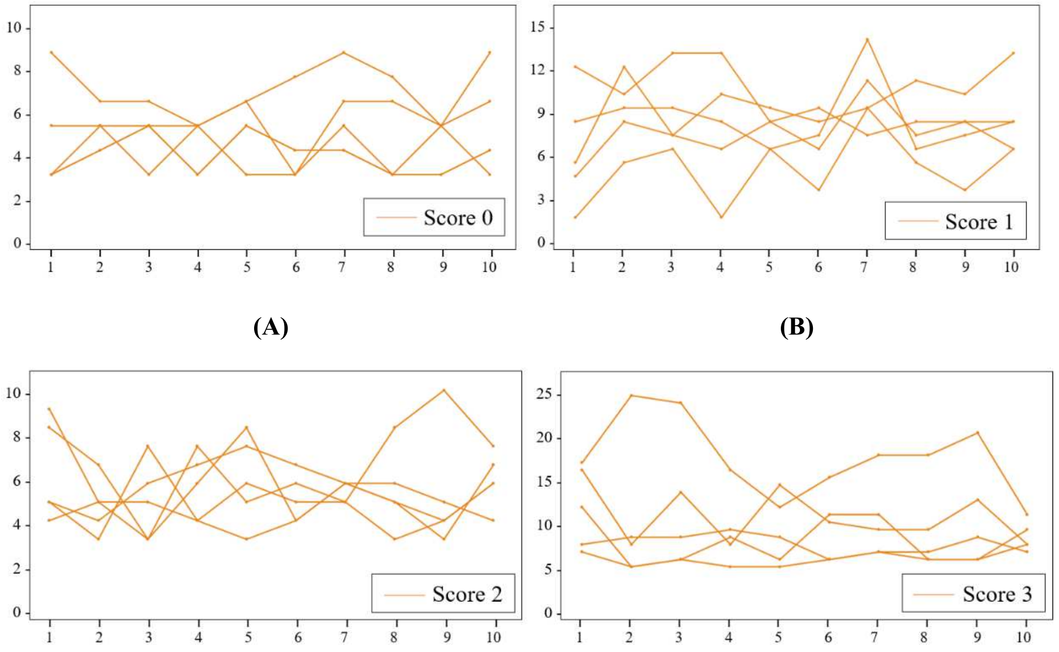
(A)

(B)

(C)

(D)

Fig. 12 Range data diagram of four fractions, the abscissa represents 10 actions, the ordinate represents the range, and the size is 0-1.



(A)

(B)

(C)

(D)

Fig. 13 Velocity data diagram of four fractions, the abscissa represents 10 actions, and the ordinate represents the velocity.

5 Conclusion

In this paper, a method combining manual features with CNN is proposed for the evaluation and classification of finger tapping in patients with PD, which can provide reliable MDS-UPDRS scores for patients with PD. The evaluation method in this paper achieves the optimal accuracy of 79.7% on the self-built dataset, which provides a new idea for PD action evaluation. And using a lot of experiments, we prove that the manual feature we designed meets the scoring criteria of MDS-UPDRS, and verifies the rationality and reliability of this method. The method without wearing additional sensor for patients has obvious advantages in remote clinical evaluation. At the same time, the manual feature we designed are suitable for small sample classification, which effectively solves the problem of difficult data acquisition in medicine. In the future, we will continue to conduct research and analysis to further verify and improve the method proposed in this paper.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request. The code used in this paper is publicly available online at

https://github.com/lizhu1126/CNN-for-PD-Action.git

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgements

This work was supported by the Zhejiang Provincial Key Lab of Equipment Electronics，Hangzhou, China. This project was also supported by the grants from Zhejiang Health New Technology Product Research and Development Project (2021PY034）, Zhejiang Medicine and Health Science and Technology Project (2021KY420) and Medical health Science and Technology project of Zhejiang (2020366842).

References

1. Tysnes O B, Storstein A. (2017). Epidemiology of Parkinson’s disease. *Journal of neural transmission*, 124(8): 901-905. https://doi.org/10.1016/S1474-4422(06)70471-9
2. More S V, Choi D K. (2016). Emerging preclinical pharmacological targets for Parkinson's disease. *Oncotarget*, 7(20):29835-29863. https://doi.org/10.18632/oncotarget.8104
3. Fang C, Lv L, Mao S, et al. (2019). Cognition Deficits in Parkinson's Disease: Mechanisms and Treatment. *Parkinson's Disease*, 2020(9):1-11. https://doi.org/10.1155/2020/2076942
4. Goetz C G, Tilley B C, Shaftman S R, et al. (2008). Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. Movement disorders: official journal of the Movement Disorder Society, 23(15): 2129-2170. https://doi.org/10.1002/mds.22340
5. XU M, CHEN T, MENG X, et al. (2020). Progress on quantitative assessments of motor symptoms for Parkinson′ s disease. *Chinese Journal of Neurology*, 845-854.
6. Wang D, Whangbo T. (2019). Automatic diagnostic system for parkinsons disease based on deep learning using midbrain magnetic resonance images. *International Journal of Advanced Science and Techn*ology, *SERSC Australia*, 124: 1-20.
7. Sivaranjini S, Sujatha C M. (2020). Deep learning based diagnosis of Parkinson’s disease using convolutional neural network. *Multimedia tools and applications*, 79(21): 15467-15479. https://doi.org/10.1007/s11042-019-7469-8
8. Mohamadzadeh S, Pasban S, Zeraatkar-Moghadam J, et al. (2021). Parkinson’s Disease Detection by Using Feature Selection and Sparse Representation. *Journal of Medical and Biological Engineering*, 2021: 1-10. https://doi.org/10.1007/s40846-021-00626-y
9. Khodakarami H, Farzanehfar P, Horne M. (2019). The use of data from the Parkinson’s KinetiGraph to identify potential candidates for device assisted therapies. *Sensors*, 19(10): 2241. https://doi.org/10.3390/s19102241
10. Andrade A, Paixo A, Cabral A M, et al. (2020). Task-Specific Tremor Quantification in a Clinical Setting for Parkinson's Disease. *Journal of Medical and Biological Engineering*, 40(6):1-30. https://doi.org/10.1007/s40846-020-00576-x
11. Chen L, Wang H, Y Huang, et al. (2020). Robust hierarchical sliding mode control of a two-wheeled self-balancing vehicle using perturbation estimation. *Mechanical systems and signal processing*, 139(May):106584.1-106584.19. https://doi.org/10.1016/j.ymssp.2019.106584
12. Parisi F, Ferrari G, Giuberti M, et al. (2016). Inertial BSN-based characterization and automatic UPDRS evaluation of the gait task of Parkinsonians. *IEEE Transactions on Affective Computing*, 7(3): 258-271. https://doi.org/10.1109/TAFFC.2016.2549533
13. Dai H, Xiong Y, Cai G, et al. (2018). A mechanical impedance-based measurement system for quantifying Parkinsonian rigidity. *Journal of biomedical engineering*, 35(3): 421-428. https://doi.org/10.7507/1001-5515.201708069
14. Zhong X, Zheng J, Ye Q. (2018). Advances in quantitative assessment of parkinsonian motor symptoms with wearable devices. *Science China Life Sciences*, 61(12): 1589-1592. https://doi.org/10.1007/s11427-018-9434-5
15. Li Y, He Z, Ye X, et al. (2019). Spatial temporal graph convolutional networks for skeleton-based dynamic hand gesture recognition. *EURASIP Journal on Image and Video Processing*, 2019(1): 1-7. https://doi.org/10.1186/s13640-019-0476-x

16. Hu G, Cui B, Yu S. (2019). Joint learning in the spatio-temporal and frequency domains for skeleton-based action recognition. *IEEE Transactions on Multimedia*, 22(9): 2207-2220. https://doi.org/10.1109/TMM.2019.2953325
17. Li C, Zhang X, Liao L, et al. (2019). Skeleton-based gesture recognition using several fully connected layers with path signature features and temporal transformer module. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01): 8585-8593. https://doi.org/10.1609/aaai.v33i01.33018585
18. Chen Y, Zhang Z, Yuan C, et al. (2021) Channel-wise Topology Refinement Graph Convolution for Skeleton-Based Action Recognition. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 13359-13368. https://arxiv.org/abs/2107.12213
19. Kay W, Carreira J, Simonyan K, et al. (2017). The kinetics human action video dataset. https://arxiv.org/abs/1705.06950
20. Shahroudy A, Liu J, Ng T T, et al. (2016). Ntu rgb+ d: A large scale dataset for 3d human activity analysis. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1010-1019.
21. Sano Y, Kandori A, Shima K, et al. (2016). Quantifying Parkinson's disease finger-tapping severity by extracting and synthesizing finger motion properties. *Medical & biological engineering & computing*, 54(6): 953-965. https://doi.org/10.1007/s11517-016-1467-z
22. Leijnse J N A L, Campbell-Kyureghyan N H, Spektor D, et al. (2008). Assessment of individual finger muscle activity in the extensor digitorum communis by surface EMG. *Journal of neurophysiology*, 100(6): 3225-3235. https://doi.org/10.1152/jn.90570.2008
23. Arias P, Robles-García V, Espinosa N, et al. (2012). Validity of the finger tapping test in Parkinson's disease, elderly and young healthy subjects: Is there a role for central fatigue?. *Clinical Neurophysiology*, 123(10): 2034-2041, 2012. https://doi.org/10.1016/j.clinph.2012.04.001
24. Foki T, Pirker W, Geißler A, et al. (2015). Finger dexterity deficits in Parkinson's disease and somatosensory cortical dysfunction. *Parkinsonism & Related Disorders*, 21(3): 259-265. https://doi.org/10.1016/j.parkreldis.2014.12.025
25. Goetz C G, Stebbins G T. (2014). Assuring interrater reliability for the UPDRS motor section: utility of the UPDRS teaching tape. *Movement Disorders*, 19(12): 1453-1456. https://doi.org/10.1002/mds.20220
26. Zhang F, Bazarevsky V, Vakunov A, et al. (2020). Mediapipe hands: On-device real-time hand tracking. arXiv preprint arXiv:2006.10214. https://arxiv.org/abs/2006.10214
27. Feichtenhofer C, Pinz A, Zisserman A. (2016). Convolutional two-stream network fusion for video action recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1933-1941.
28. Kalfaoglu M E, Kalkan S, Alatan A A. (2020). Late temporal modeling in 3d cnn architectures with bert for action recognition. *European Conference on Computer Vision. Springer, Cham*, 2020: 731-747. https://doi.org/10.1007/978-3-030-68238-5_48
29. Shi L, Zhang Y, Cheng J, et al. (2019). Two-stream adaptive graph convolutional networks for skeleton-based action recognition. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 12026-12035.
30. Wen Y H, Gao L, Fu H, et al. (2019). Graph CNNs with motif and variable temporal block for skeleton-based action recognition. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01): 8989-8996. https://doi.org/10.1609/aaai.v33i01.33018989

31. Li H, Shao X, Zhang C, et al. (2021). Automated assessment of Parkinsonian finger-tapping tests through a vision-based fine-grained classification model. *Neurocomputing*, 441: 260-271. https://doi.org/10.1016/j.neucom.2021.02.011
32. Wang Z, Yan W, Oates T. (2017). Time series classification from scratch with deep neural networks: A strong baseline. *2017 International joint conference on neural networks (IJCNN). IEEE*, 1578-1585. https://doi.org/10.1109/IJCNN.2017.7966039
33. Rakthanmanon T, Campana B, Mueen A, et al. (2012). Searching and mining trillions of time series subsequences under dynamic time warping. *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 262-270. https://doi.org/10.1145/2339530.2339576