Medical Image Analysis 85 (2023) 102754

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Vision-based estimation of MDS-UPDRS scores for quantifying Parkinson's disease tremor severity

Weiping Liu [a,d,1], Xiaozhen Lin [b,1], Xinghong Chen [a,d], Qing Wang [a,d], Xiumei Wang [a,d], Bin Yang [a,d], Naiqing Cai [c], Rong Chen [a,d], Guannan Chen [a,d,*], Yu Lin [c,**]

[a] Key Laboratory of OptoElectronic Science and Technology for Medicine of Ministry of Education, Fujian Normal University, Fuzhou 350007, China
[b] Department of Geriatrics, The First Affiliated Hospital, Fujian Medical University, Fuzhou 350005, China
[c] Department of Neurology and Institute of Neurology, The First Affiliated Hospital, Fujian Medical University, Fuzhou 350005, China
[d] Fujian Provincial Key Laboratory of Photonics Technology, Fujian Normal University, Fuzhou 350007, China

ARTICLE INFO

Keywords:
Parkinson's disease
Rest tremors
Postural tremors
EVM pre-processing
Global Temporal-difference Shift Network

ABSTRACT

Parkinson's disease (PD) is a common neurodegenerative movement disorder among older individuals. As one of the typical symptoms of PD, tremor is a critical reference in the PD assessment. A widely accepted clinical approach to assessing tremors in PD is based on part III of the Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS). However, expert assessment of tremor is a time-consuming and laborious process that poses considerable challenges to the medical evaluation of PD. In this paper, we proposed a novel model, Global Temporal-difference Shift Network (GTSN), to estimate the MDS-UPDRS score of PD tremors based on video. The PD tremor videos were scored according to the majority vote of multiple raters. We used Eulerian Video Magnification (EVM) pre-processing to enhance the representations of subtle PD tremors in the videos. To make the model better focus on the tremors in the video, we proposed a special temporal difference module, which stacks the current optical flow to the result of inter-frame difference. The prediction scores were obtained from the Residual Networks (ResNet) embedded with a novel module, the Global Shift Module (GSM), which allowed the features of the current segment to include the global segment features. We carried out independent experiments using PD tremor videos of different body parts based on the scoring content of the MDS-UPDRS. On a fairly large dataset, our method achieved an accuracy of 90.6% for hands with rest tremors, 85.9% for tremors in the leg, and 89.0% for the jaw. An accuracy of 84.9% was obtained for postural tremors. Our study demonstrated the effectiveness of computer-assisted assessment for PD tremors based on video analysis. The latest version of the code is available at https://github.com/199507284711/PD-GTSN.

1. Introduction

Parkinson's disease (PD) is a neurodegenerative disease with a high incidence in older individuals (Bhattacharjee and Sambamoorthi, 2013; Dong et al., 2021). It has a complex pathogenesis and can result in severe deterioration in the patient's health. As the ageing population increases, the number of PD patients is on the rise, which puts much pressure on diagnostic process in healthcare settings (Amoroso et al., 2018; Porritt et al., 2006). Therefore, a computer-assisted assessment method is urgently needed so that specialists can develop a treatment plan early. The symptoms of PD include tremors, bradykinesia, rigidity, and postural instability (Bi et al., 2021). Tremor is a kind of abnormal movement defined as the involuntary occurrence of periodic oscillations of body parts, which has been suggested to occur in more than 70% of PD patients (Baumann, 2012; Politis et al., 2010). Thus, the accuracy of PD tremor assessment is essential for PD diagnoses and treatment (Bhatia et al., 2018; Massano and Bhatia, 2012). However, accurately assessing PD tremors is a major challenge since the form and amplitude of PD tremors often vary and are context-dependent (Zach et al., 2015).

The current clinical assessment of PD is primarily based on the

* Corresponding author at: Key Laboratory of OptoElectronic Science and Technology for Medicine of Ministry of Education, Fujian Normal University, Fuzhou 350007, China.
* Corresponding author at: Department of Neurology and Institute of Neurology, The First Affiliated Hospital, Fujian Medical University, Fuzhou 350005, China
E-mail addresses: edado@fjnu.edu.cn (G. Chen), yulinwin2009@aliyun.com (Y. Lin).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.media.2023.102754
Received 16 May 2022; Received in revised form 7 January 2023; Accepted 18 January 2023
Available online 20 January 2023

Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS), which was defined by the World Academy of Movement Disorders (Bhatia et al., 2018). It has four parts with several tests for various motors such as gait, finger tapping, and leg agility. Each test is scored by trained experts using five levels of severity: 0=normal, 1=slight, 2=mild, 3=moderate, and 4=severe (Zitser et al., 2017). Part III of the MDS-UPDRS contains the tremor tests for PD patients, with the 3.17 test for rest tremor and the 3.15 test for postural tremor (Li et al., 2018c). Fig. 1 describes the tremor test in MDS-UPDRS. The 3.17 test for rest tremor requires participants to sit statically in a chair for 10 s with their hands resting on the chair's armrests and their feet placed comfortably on the floor. The specialist scores the participant by observing the limb and jaw tremors. In the 3.15 test for postural tremors, the participant holds their arm straight out in front of the body with the palm down. The wrist is held in a straight position, and the fingers are separated. Then the specialist observes the participant's hand tremors to give the score (Goetz et al., 2012; Stebbins et al., 2013). The MDS-UPDRS provides a reliable criterion for the assessment of PD severity. However, the evaluation process requires trained experts to complete, which results in inefficiencies in healthcare as well as subjective variation in assessment. In addition, the limitations to mobility and travel restrictions caused by the recent COVID-19 pandemic make it more challenging for PD patients to get timely clinical assessment and treatment (Li et al., 2021). Therefore, clinical practice and health services urgently need an automated and objective method to assess PD tremors. We argue it is feasible to quantify the severity of PD tremors based on video by computer-assisted technologies.

Thus far, most research on automatic assessment of PD tremors has focused on wearable sensors (Monje et al., 2019; Silva de Lima et al., 2017). Inertial sensing-based wearable devices (ISWDs) were designed to capture tremors, and the scores of professional physicians were trained by a supervised learning algorithm. However, the results of this method have not been satisfactory. Some researchers have combined the 6-axis high-precision electromagnetic tracking system (EMTS) with machine learning algorithms to address this challenge, which has improved the quantification performance (Dai et al., 2021). It should be noted that the wearable device might affect the patient's tremor performance during wearing due to the weight of the device, which could cause the results to be less accurate. Moreover, wearing the device might interfere with the patient's daily life. Thus, the sensor method is not applicable for daily monitoring. In this study, we developed an automated video-based method for assessing the severity of PD tremors. This novel method did not require the patient to put on any wearable equipment, which would not cause any adverse effects on the patients.

Base on the idea of a video classification algorithm, we proposed a deep learning network model to quantify PD tremors according to the MDS-UPDRS. One challenge to the video classification algorithm was the lack of video datasets. Deep learning algorithms often need numerous datasets to sufficiently support the training task of the network, so a data augmentation process was necessary (Wang et al., 2021a, 2022, 2021b). In addition, the scoring targets for PD tremors in the MDS-UPDRS involved several different body parts (Legaria-Santiago et al., 2022). This would have required considerable workload to capture the videos individually. So it is necessary to design a method to divide video of different body parts from the entire video automatically. Human pose estimation predicts the coordinates of the body joints in the image using a deep learning algorithm. Subsequently, the body joints are connected based on the structure of the human skeleton (Zhao et al., 2021). We postulated that we could accurately segment the entire video through the coordinates that were obtained from the pose estimation. Based on this process, it was possible to obtain the video dataset of target body parts automatically.

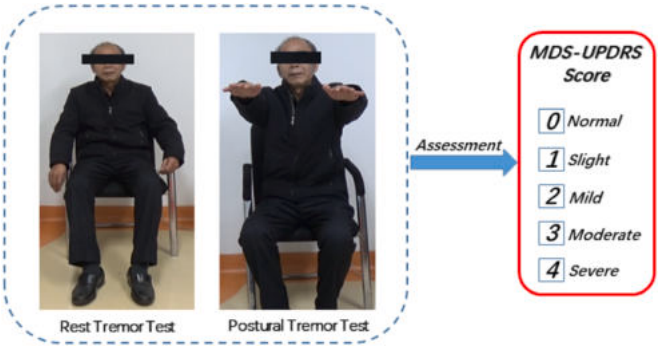

Fig. 1. Rest and postural tremor test in MDS-UPDRS for PD patients.

Recording PD tremors using video has the potential problem that very subtle tremors might not be adequately represented in the video, skewing the final results. To solve this challenge, we adopted the technique of video magnification to enhance the subtle tremors in the video, thereby increasing the accuracy of the results. Apart from that, in video classification tasks based on deep learning, a large amount of input data and complex network models would result in substantial calculations (Wang et al., 2020). Therefore, the design of the video sampling and the network structure were particularly significant. Current video classification algorithms aim to recognize larger actions with greater sampling intervals (Zhou et al., 2017). However, PD tremors were more minute motions, so the sampling interval should be appropriate. Since too small sampling interval would increase the occurrence of unnecessary computations, while an over large sampling interval would decrease the accuracy of the results. Therefore, the best sampling method for the video also was a focus of our research. Moreover, current video-based methods are only concerned with local temporal movements, while assessing PD tremors is a global process. Thus, our method expanded the model's focus to include the whole temporal domain of the videos.

In summary, we proposed an effective method to predict the MDS-UPDRS scores of PD tremor videos in this paper. The contribution of our work is as follows:

(1) We proposed an efficient video-based method to evaluate the severity of PD tremors, which was carried out without physical contact with the patient. This novel method has achieved considerable accuracy in predicting MDS-UPDRS scores for both rest and postural tremors.
(2) We combined temporal difference and video classification algorithms to classify PD tremor videos with different severity levels effectively. EVM (Eulerian Video Magnification) was applied as video pre-processing to enhance the subtle PD tremors in the videos.
(3) We proposed the GTSN model, which focused more on global temporal changes in the PD tremor videos. This novel model is more advanced than other video classification models to assess PD tremors in videos.
(4) We proposed a novel module, GSM, which allowed each temporal segment to include the feature of the global segment. Experimental results demonstrated that GSM was more powerful than the other state-of-the-art modules in predicting the scores based on the PD tremor videos.

2. Related work

Researchers have recently invented several methods to assess the severity of PD. These researchers have primarily focused their work on analyzing the gait and bradykinesia of PD patients using specially designed wearable sensors (Fino and Mancini, 2020; Liu et al., 2019). With the advent of deep learning algorithms, methods based on videos have been proposed to quantify PD (Hughes et al., 2020; Lu et al., 2021; Mei et al., 2021). Although the final results of these methods have proven to be satisfactory in experiments, the application of these methods in the clinical diagnosis of PD has not yet been adopted.

Mandy Lu introduced an ordinal focal neural network and used human joint information as inputs to the model, which achieved acceptable predictions of MDS-UPDRS scores for PD gait. They also applied this method to the finger-tapping task in the MDS-UPDRS and obtained excellent results (Lu et al., 2020, 2021). Besides taking the video directly to assess the severity of PD, some works extracted motion features including speed and frequency from the video for further analysis. They extract the features that were most relevant to the motion and carried out the assessment of PD using deep learning (Vignoud et al., 2022). Samuel Rupprechter calculated six motion features, including speed, arm swing, postural control, and smoothness, using a sequence of key points that were obtained from videos. Subsequently, they trained an ordinal random forest classification model, which produced a rating estimation of the MDS-UPDRS gait on a lager video dataset (Rupprechter et al., 2021).

Video-based deep learning algorithms have been used to assess MDS-UPDRS as well as other assessment methods. Michael H. Li was the first to apply deep learning algorithms for the vision-based assessment of PD and levodopa-induced dyskinesia (LID). They used the Random Forest algorithm to predict the total UPDRS and MDS-UPDRS Part III scores based on motion features (Li et al., 2018a, b). In addition, some research has combined video-based methods with other methods to evaluate the results in an integrated manner. They used special sensors to obtain 3D gait information. Then, they combine it with 2D gait information from the camera to demonstrate the correlation between these two methods and the MDS-UPDRS-gait and SAS-gait scores (Sabo et al., 2020). Vincenzo Dentamaro extended the application of the video-based gait assessment method to other diseases. In their work, the human gait movement patterns were modeled using the kinematic theory of rapid human movements and its sigma-lognormal model to achieve diagnoses for neurodegenerative diseases (Dentamaro et al., 2020). With the development of human posture estimation algorithms and graph neural networks, researchers take human key points as input of graph neural networks to analyzed the characteristic relationships between movements. Rui Guo proposed a sparse adaptive graph convolutional network (SA-GCN) in their study to achieve a fine-grained quantitative evaluation of skeleton sequences that had been extracted from videos. They demonstrated the validity and reliability of their method for PD motor disorder assessment using a large dataset (Guo et al., 2020).

Wearable watches and smartphones also have been used in PD tremor research (Liddle et al., 2014; Marxreiter et al., 2020). Kuosmanen et al. quantified the severity of PD hand tremors using smartphone inertial sensors to better understand the effects of PD medications in normal environments. These studies have demonstrated the effectiveness of the method in monitoring PD symptoms and remotely assessing the effects of medications (Kuosmanen et al., 2020). It should be noted that wearable devices could influence the patient's tremor performance due to the weight of the device, which might adversely affect the accuracy of the results. Moreover, wearable devices are expensive, complex, and not easy to design, which is more inconvenient than the non-intrusive method based on video.

Feature extraction algorithms of videos based on deep learning are better at quantifying PD tremors captured on video (Ali et al., 2020; Rupprechter et al., 2021), but previous research has focused on only a few video features. On the other hand, the quantification process can be considered as a process of video classification since the score of videos (0,1,2,3,4) is discrete. Zhao Yin employed a 3D Convolutional Neural Network (CNN) to quantify the severity of PD, which was processed using PD videos. Due to the limited video data, they pre-trained the network model using a non-medical dataset. The authors applied transfer learning in their research, and the experimental results proved to be valid. However, their study only targeted the seven tasks in MDS-UPDRS and ignored the rest tremors test in part III of the MDS-UPDRS (Yin et al., 2022). Haozheng Zhang proposed a new model named SPAPNet, which classified tremors using non-invasive video recordings. They extracted relevant tremor information and effectively filtered out noise using a novel attention module with a lightweight pyramidal channel squeeze fusion architecture (Zhang et al., 2022). Their work focused on classifying different tremor types, while our work was about MDS-UPDRS assessment of PD tremors which was crucial for following the progression of patients' symptoms.

Silvia L. Pintea et al. accomplished the task of estimating hand tremor frequencies from RGB videos by proposing the Eulerian method, which used intensity values or phase information from the video to estimate the result after removing large motions from the videos (Pintea et al., 2018). This method developed a novel concept for research through the analysis and diagnosis of dyskinesia, such as PD. Finally, Xinyi Wang designed a gesture recognition and body motion detection system. In this system, relevant features were extracted from videos taken in arbitrary situations, and machine learning was used to make classifications based on the observed video features to detect tremors (Wang et al., 2021c).

3. Materials and methods

3.1. Participants and dataset

In this study, 130 unrelated sporadic patients were recruited from 2019 to 2021. The diagnosis of PD was based on the UK Brain Bank diagnostic criteria, which was proposed by The United Kingdom Parkinson's Disease Society Brain Bank (UKPDSBB). This system is the first set of formal diagnostic criteria for PD, which is currently broadly used in clinical trials and routine clinical practice worldwide (Luca et al., 2018). This study was approved by the Ethics Committee of the First Affiliated Hospital of Fujian Medical University, and written consent was obtained from all participants. All participants in this paper were evaluated using the MDS-UPDRS in the off-medication state, and the evaluation process was recorded on video. These videos were produced and provided by professional specialists who participated in the International Movement Disorders Association's MDS-UPDRS scoring training and received a certificate of competency. The videos were not allowed to be made public according to the relevant regulations protecting patient privacy. All video data were recorded using one Sony camera with a video resolution of 1920×1280, a video frame rate of 30 frames per second (fps), and the video recordings were stored in MTS format. Each participant was captured on two videos with the guidance of professionals. One video was obtained for the 3.17 test, and the other for the 3.15 test. During the recording process, the camera was positioned 3 m away from the participants and kept in a stationary position. In the videos for the 3.17 test, the participants were instructed to sit calmly in a chair for 7 to 14 s with their hands on the arms of the chair and their feet resting comfortably on the floor. The videos for the 3.15 test required the participants to sit in a chair with their arms held straight out from their body, palms facing down, and wrists straight while keeping their fingers apart.

According to the MDS-UPDRS, the test for rest tremors (3.17) focused on three items: hands, legs, and jaw. The test for postural tremors (3.15) focused only on the hands. Therefore, we focused on scoring these three body parts in our study. The scores from the videos were evaluated by three trained, board-certified physicians. The final scores were determined based on a majority vote of the three evaluators. Since the scarcity of videos with scores of 3 and 4 in the collected videos, we combined them with videos scoring 2 as scored 2+. Table 1 and 2 shows the valid number of different body parts of the participants that were assessed as well as the gender distribution. Except for the jaw assessment, the videos of the hands and legs included the left and right sides. Therefore, twice the amount of leg and hand video data were obtained. We also augmented the videos by mirroring to expand the dataset for each body part examined. Thus, we theoretically obtained four times as many videos of the hands and legs as the number of participants and twice as many videos of the jaw as the number of participants. However, the automatic extraction of the different body parts included some problems

Table 1
Dataset of rest tremor participants from 130 participants in this study. The ground-truth for each video is determined by the majority vote of raters. (M: Male; F: Female)

Score	Left Hand Valid/Total	M/F	Right Hand Valid/Total	M/F	Left Leg Valid/Total	M/F	Right Leg Valid/Total	M/F	Jaw Valid/Total	M/F
0	46/47	28/19	46/49	22/27	41/49	28/21	48/56	31/25	58/61	27/34
1	40/41	20/21	34/36	18/18	55/56	24/32	45/46	22/24	40/43	22/21
2	29/29	14/15	35/35	20/15	22/24	15/9	23/26	13/13	23/25	18/7
3	11/11	4/7	8/9	6/3	0/1	0/1	2/2	1/1	1/1	0/1
4	2/2	1/1	1/1	1/0	0/0	0/0	0/0	0/0	0/0	0/0
All	128/130	67/63	124/130	67/63	118/130	67/63	118/130	67/63	122/130	67/63

e.g. if one hand of a patient occluded the other one, or some participants' voluntary head bowing and shaking. All the above non-standard movements would cause failed acquisition for body parts video, which also leads to the number of available videos unequal to the total number of videos. The actual number of valid video of different body parts in our experiment is shown in Table 3.

3.2. Overall framework

Our method contained EVM (Eulerian Video Magnification) pre-processing and GTSN (Global Temporal-difference Shift Network), which included the temporal difference module and the ResNet-50 embedded with GSM (Global Shift Module). The overall flow is shown in Fig. 2. First, the input tremor video was pre-processed by EVM. Then after the sampling process, the optical flow was extracted from the sampled image as the input of GTSN. In the GTSN, the optical flow initially passed through the temporal difference module, stacking the inter-frame difference results of five successive frames with the current frame. Then the features were extracted from the ResNet-50 with GSM. The GSM was placed inside the residual branch in a residual block, which was denoted as the GSM-Residual block. This plug-and-play module allowed the features of each moment to accept the features of the other moments through the global temporal shift. Finally, the predicted score of the video was output from a fully connected layer. In the following sections, we described the proposed method in detail.

3.3. Obtaining videos of different body parts

The objects scored in tests 3.17 and 3.15 of the MDS-UPDRS included the participant's limbs and jaw. We used OpenPose to automatically divide the original video to obtain partial videos containing only the target body parts (Cao et al., 2021). This procedure greatly reduces the workload of collating tremor videos from different body parts. 21 hand key points of OpenPose were used to obtain the videos which only contain the hands; 25 body key points of OpenPose were used to obtain the videos which only contain the legs; 20 mouth key points of OpenPose were used to obtain the videos which only contain the jaw. The distribution of OpenPose keypoints and the acquisition of the video are shown in Fig. 3.

3.4. Video magnification

Eulerian Video Magnification has been proposed by Wu, HY. It focuses on subtle motions in videos that can be effectively enhanced through filtering and amplification processes (Wu et al., 2012). In our research, the three raters stand at the same angle and distance from the camera during the PD tremor video recording. Considering that the resolution of the human eyes is greater than that of the common camera we used in our research, it may be difficult for some subtle tremors to be captured and presented in the videos. Therefore, we pre-processed the PD tremor video using EVM to amplify tremors that were difficult to be

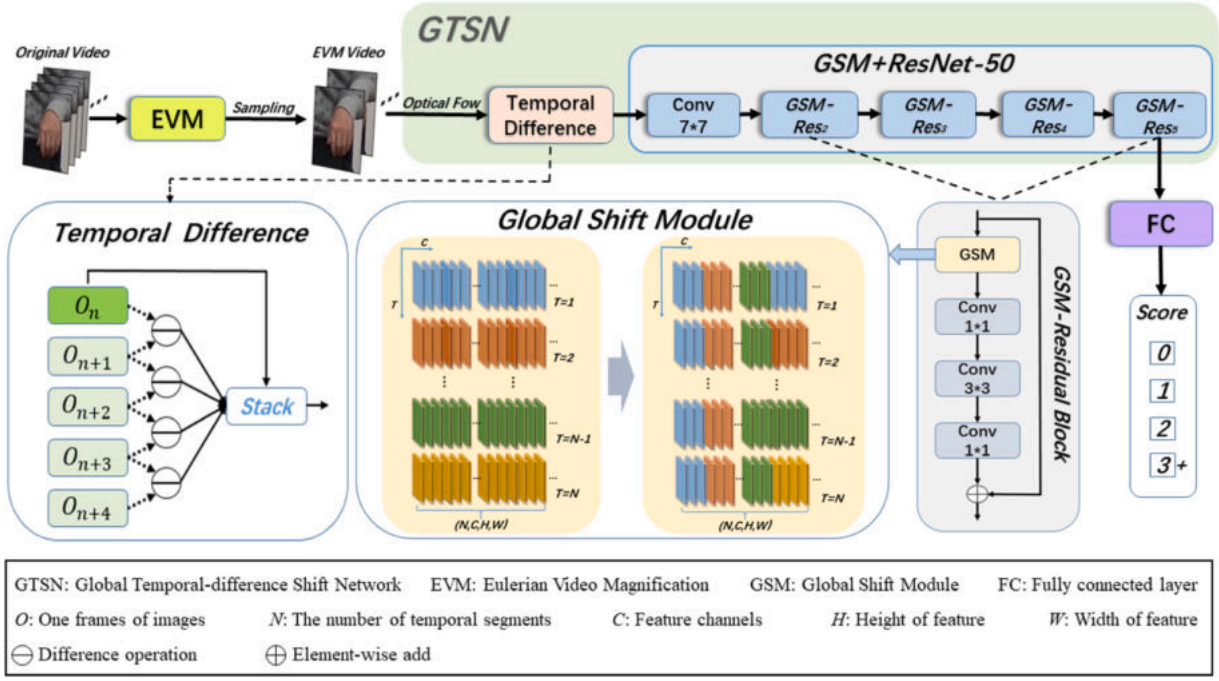

Fig. 2. The framework of our method: the subtle tremor in the video is amplified by EVM pre-processing. In the GTSN, temporal difference module stacks the current optical flow to the result of inter-frame difference. The features are obtained through the ResNet-50 which consists of several GSM-Residual block. Each GSM-Residual block is embedded with GSM to gain the global temporal features. The final prediction score is obtained through a fully connected layer.

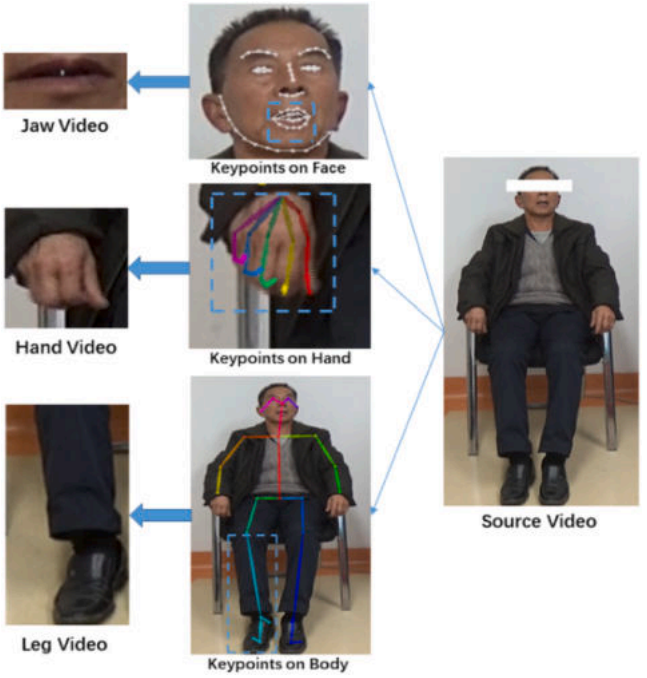

Fig. 3. Videos of different body parts are obtained from the source video by the keypoints of the OpenPose. The lines in the middle figures represent the human skeleton, and different colors mean different parts. The videos containing only the target body parts can be automatically obtained through OpenPose.

picked in the video. For EVM, a video *I*can be indicated as amplitude and phase:

$$I(m,n,t) = A(\gamma,\theta,m,n,t)e^{i\psi(\gamma,\theta,m,n,t)} \tag{1}$$

Where γ represented the scale and θ represented the direction. The phase-changed δ at time t was calculated based on the phase at the time t_0:

$$\delta(\gamma,\theta,m,n,t) = \psi(\gamma,\theta,m,n,t) - \psi(\gamma,\theta,m,n,t_0) \tag{2}$$

After obtaining the phase change, we multiplied it by the amplification factor α and added it to the phase of the original video to obtain the magnified result:

$$\widehat{I}(m,n,t) = A(\gamma,\theta,m,n,t)e^{i(\psi+\alpha\delta)} \tag{3}$$

Where $\widehat{I}$ was the video that has been magnified. We used EVM as a preprocessing technique to enhance subtle movements in the original video. We apply EVM to all PD tremor videos and compare the result from the original video and that from the EVM video, which shows a significant difference for all PD tremor videos. To visualize the differences between the original video and the EVM video, we took pixel slices of the same position and spliced them together. Fig. 4 shows the splicing results of slices from one original video and its EVM video. This video was a rest tremor video of one participant's right hand. Comparing the two splicing results of the slices on the right side of Fig. 4, we determined that the EVM effectively magnified the subtle movements in the original video.

3.5. Global temporal-difference shift network

In the algorithms of video classification based on the two-stream network, the inputs were of two types: RGB images and optical flow (Lee et al., 2018; Wang et al., 2019). RGB images have three channels with the input shape of $N{\times}3{\times}c{\times}H{\times}W$, and optical flow has two channels with the input shape of $N{\times}2{\times}c{\times}H{\times}W$. In these cases, N was the number of video segments, and c was the number of each segment

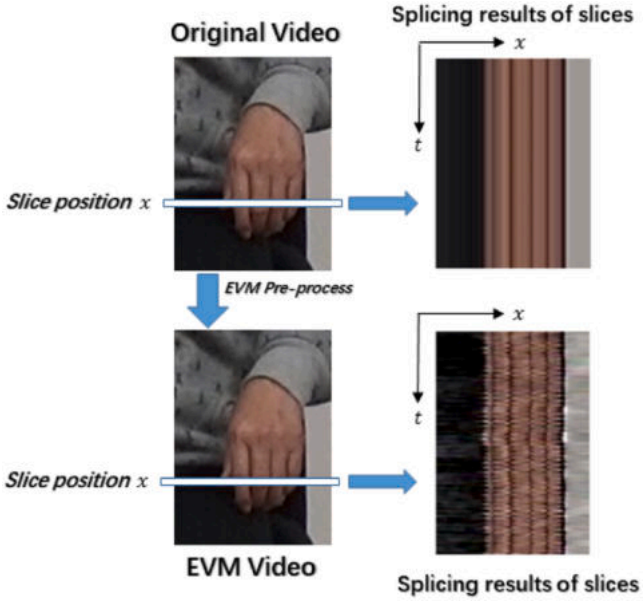

Fig. 4. The splicing slices of the original video and the video after EVM in the same position. The splicing results of slices showed that EVM can enhance the subtle PD tremors in the video. Top left: we take a pixels slice from each frame of the original video at row *x*; Top right: we splice the original video slices in the order of temporal *t* to show the pixels' motion at the slice position in the original video; Bottom left: we take a pixel slice from each frame of the EVM video at row *x*; Bottom right: we splice the EVM video slices in the order of time *t* to show the pixels' motion at the slice position in the EVM video.

channel. Current video classification methods tend to target larger actions. For the more subtle movements of PD tremors, we proposed a new network model GTSN.

3.5.1. Video sampling

To minimize the number of calculations in the video classification algorithm, the videos must be appropriately sampled as input to the model (Liu et al., 2021b). The method proposed in this study focused on the PD tremor videos. We argue that the process of recording PD tremors with a camera was a sampling of the PD tremors, where the fps of video were the sampling frequency (e.g., a video at 30 fps that had a sampling frequency of 30 Hz). Therefore, the actual tremor frequency and video sampling frequency of PD patients needed to satisfy the Nyquist limits to ensure the validity of the results. Thus, the frames per second of the captured video *FPS* and the highest frequency of the PD tremor $f_{\max}$ needed to satisfy the following inequalities:

$$FPS > 2{\cdot}f_{\max} \tag{4}$$

Related studies have shown that the normal frequency range of PD tremors is 3 to 7 Hz (Delval et al., 2016; Duval and Beuter, 1998; Florin et al., 2008; Lukhanina et al., 2000). It was apparent that the video in our study obeyed the Nyquist limits. Rest and postural tremors commonly appear after the PD patient has been in a stable state for a while. The tremor might not occur immediately in PD patients during the first half of the video in our research. Therefore, we only captured frames to 3 s later as the model's input. Video classification algorithms tend to use sparse sampling or dense sampling (Liu et al., 2018; Vig et al., 2012). However, both methods are unsuitable for the tremor video because the tremor is a continuous and small motion. Therefore, we especially designed a sampling method for PD tremor video in this study. Furthermore, to reduce the number of calculations as much as possible without distortion, we sampled the video in compliance with the Nyquist limit:

$$FPS' > 2{\cdot}f_{\max} \tag{5}$$

Where f_{max} was set to the maximum value of 7, so the minimum sampling frequency $FPS^{'}$ could be set to 15. The original video frame rate was 30 fps, and we sampled each frame at intervals, so the final sampling result was 15 frames per second, satisfying the above inequality. Fig. 5 shows how we sampled the frames, which were only *1/2* of the original number of frames.

3.5.2. Temporal difference

Although the video classification base on deep learning can effectively classify different videos, it primarily addresses videos with larger actions (Lin et al., 2020). However, the tremor videos in our works are more subtle and slight motion, and we argue that the small differences between frames in the tremor video were particularly significant. Therefore, we designed a temporal difference module, which allowed the model to extract motion changes in the videos more effectively. We set each segment S containing five frames of images O:

$$S = [O_n, O_{n+1}, O_{n+2}, O_{n+3}, O_{n+4}] \tag{6}$$

The shape of O was $(H \times W)$. To obtain the distinctive features between consecutive frames in the tremor video, we first made an inter-frame difference on five consecutive frames in each segment and then stacked the local frame to ensure the information of the current moment could be acquired. Using the results of the temporal difference as input to the model allowed the model to focus more on the changes between frames, which might increase the model's accuracy for tremor classification. Moreover, the input included the local frame, so the model did not lose local information. To keep the number of sampled frames within the original number of frames, the start of sampling was limited to the first 12 frames. Therefore, the first segment $S_1^{'}$ after the change was:

$$S^{'}{}_1 = [O_n, O_{n+1} - O_n, O_{n+1} - O_{n+1}, O_{n+3} - O_{n+2}, O_{n+4} - O_{n+3}](0 \le n \le 12) \tag{7}$$

We took five frames from each segment as one feature channel of the model input. After the above processing, the input of shape $(N \times c' \times 5 \times H \times W)$ was kept constant, where $c'=3 \times c$ when the input was RGB image and $c'=2 \times c$ when the input was optical flow.

3.5.3. Global shift module

The Temporal Shift Module (TSM) was proposed by Ji Lin et al. It allowed the model to learn the different information between adjacent segments by shifting the feature channels up and down in the temporal dimension. Through this simple module, each segment contains features of the adjacent segments without any extra calculation (Liu et al., 2021a). However, we determined that the scoring of PD tremors by the evaluators was a global procedure. Thus, not only the adjacent segments but also other non-adjacent segments were significant for the current segment. We developed a novel plug-and-play module, GSM, which allowed for better interactions between each feature across the global temporal.

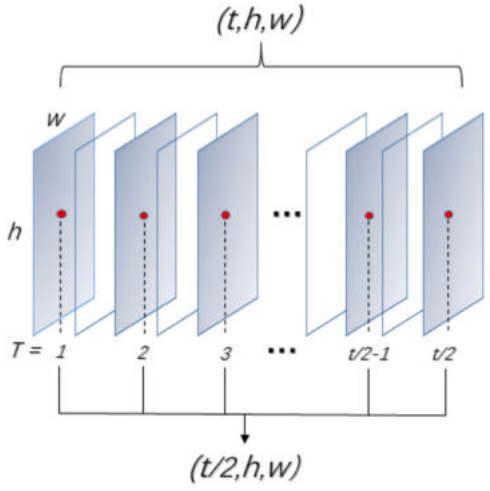

Fig. 5. By sampling the video in one frame intervals, the input shape of the model has changed from *(t, h, w)* to *(t/2, h, w)*, which reduces the data size while satisfying the Nyquist limits.

We extended the shifted feature channels to all temporal segments so that each segment incorporated the temporal difference information from the others. Thus, the global tremor information of the video could be learned. Fig. 6 shows the structure of GSM, where we retained a certain percentage of the feature channels and recombined others. After the global temporal shift, the current moment segment still retained most of the original features, while the global features represented only a small proportion. GSM allowed the information of every temporal feature to be exchanged with each other sufficiently so that the global features of the PD tremor could be learned, which increased the sensitivity of the convolutional neural network to the whole tremor video. Notably, the shifting object of the TSM was the features of the frames in the temporal segment, whereas GSM dealt with the features of the current frames and the temporal difference.

For input shape $(N \times C \times H \times W)$, N was the number of temporal segments, and C was the total number of feature channels. We set the proportional size of the global shifted features to *1/m*. The shape of the features taking part in the global shift was $(N \times C/m \times H \times W)$, and the shape of the retained features was $(N \times C(m\text{-}1)/m \times H \times W)$. Each shifted feature size was $(1 \times C/(m \times N) \times H \times W)$, and the overall feature shape remained the same after the global temporal shift.

As with TSM, GSM is a plug-and-play module (Voillemin et al., 2021). It needs to be embedded in other networks. In this study, we used ResNet-50 as the backbone of our model because we found that ResNet-50 works best through experiments. We tried several ways to insert GSM into the residual block, and the version shown as GSM-Residual block in Fig. 2 demonstrates the best result. Furthermore, for the proportional size of the global shifted features *1/m*, we determined that the best performance of the model was achieved when it was set as *1/4,* which was consistent with the experimental results of TSM.

3.5.4. GTSN-loss

Formally, for K segments of the input video $(S_1, S_2, \cdots, S_K)$, each segment contained the same number of frames. By modeling the network through which the input passes, we obtained the following results:

$$\mathrm{GTSN}(S_1, S_2, \cdots, S_K) = Softmax(F(W(S_1), W(S_2), \cdots, W(S_K))) \tag{8}$$

Where the W function was the selected convolutional network (ResNet-50), and its result was the output of the convolutional neural network (ResNet-50). The F function was the feature fusion function, which fused the output feature results of all segments. Finally, the probability of the input video on every classification was derived by the *Softmax* function, which resulted in an M-dimensional vector, where M was the number of classifications of the video. Based on the standard cross-entropy loss function, the final loss function of GTSN obtained was as follows:

$$Loss(y,f) = -\sum_{i=1}^{M} y_i \left(f_i - \log \sum_{j=1}^{M} \exp f_j \right) \tag{9}$$

Where y is the true label of the input classification and f is the result of feature fusion:

$$f = F(W(S_1), W(S_2), \cdots, W(S_K)) \tag{10}$$

4. Experiments and results

In this section, we present the experimental results of the rest tremor and the postural tremor tests. We completed several comparative experiments under the combination of different modules and showed the

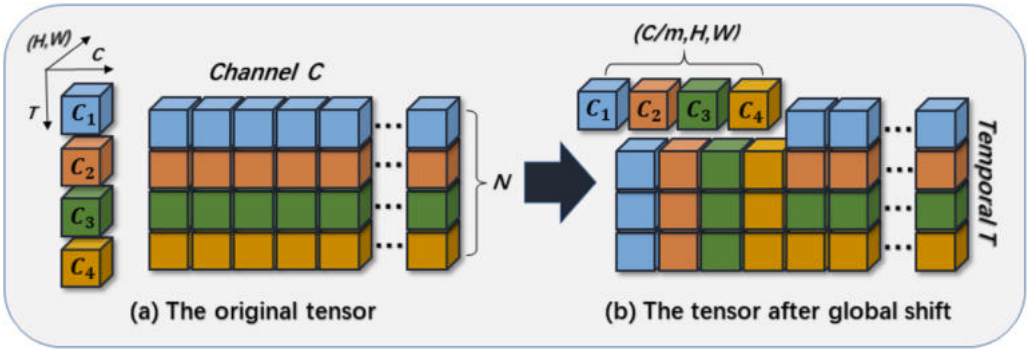

Fig. 6. Global Shift Module: Through global temporal shift, each temporal segment includes the features of other moments while retaining the majority of itself.

results of the evaluation metrics. We separated the videos for the different body parts. All videos were divided into four groups for training: left and right hand videos in the 3.17 test for hand rest tremors; left and right leg videos in the 3.17 test for leg rest tremors; jaw videos in the 3.17 test for jaw rest tremors; left and right hand videos in the 3.15 test for postural tremors. It should be noted that we trained the left and right limb videos of one participant together because the tremors of the left and right limbs from the same participant were not correlated (e.g. some participants presented a normal score of 0 for the left hand but a severe score of 2 for the right hand).

4.1. Experiment settings and evaluation metrics

All experiments in this section were run on Pytorch in a server with 4 Nvidia GTX1080ti GPU. The model used a cross-entropy loss function and the SGD optimizer. The best performance was achieved at 100 epochs with an initial learning rate of 0.002 (decays by 0.1 at epochs 40&80), a batch size of 32, and a dropout of 0.3. The backbone of the model was ResNet-50. For GSM, we embedded it in the residual block of ResNet-50. The proportional size of the global shift m in GSM was set to $1/4$. The number of video segments N was set to 8. The training time in all experiments was approximately 1 to 2 h.

To evaluate the experimental results of our method, we reported four metrics for each experimental result: average F_1, the area under the ROC curve (AUC), precision (Pre), and recall (Rec). These metrics proved the effectiveness of the method in the classification task (Ong et al., 2012; Ureten and Maras, 2022). In this study, we compared these metrics results for each score. Furthermore, we also compared the result of multiple classification and binary classification (scores 0 and non-0). We must emphasize that participants with a task score of 0 did not mean that they were not PD patients, but rather indicated that the participants might have a lower severity on this test. To account for the limited dataset size, all evaluations of this research were performed using a participant-based k-fold cross-validation with $k = 5$. The training set and test set had the same distribution. All metrics in this section were obtained from the experimental results of the test set. The videos of the test and training sets were both taken from different participants to ensure the model's generalizability.

4.2. Rest tremor score estimation results

The rest tremor experiment included rest tremors of the hands, legs, and jaw. We analyzed the results of the average F_1, the area under the ROC curve (AUC), precision (Pre), and re-call (Rec) for the experiment on the three body parts. We used the experiment of hand rest tremor videos for the representative comparison since the hand rest tremor videos were more sufficient and the scores were more uniformly distributed. We analyzed the experimental results of the hand rest tremors in detail to further explain the reasons for discrepancies between the different methods.

Table 4 shows the accuracy of the different methods in normal and EVM pre-processed cases. These models are current state-of-the-art methods in video classification tasks (Carreira and Zisserman, 2017; Feichtenhofer, 2020; Feichtenhofer et al., 2019). These results included the comparison between the RGB image and the optical flow of PD tremor videos. Since some models focused only on the RGB images, we could only show their RGB results in Table 4. In the experiments for these methods, we only changed the sampling methods to satisfy the Nyquist limits. In addition, we added the temporal difference module to compare with our method. The best results of every experiment were obtained by adjusting the hyperparameters. Every method also used the same training and test sets. All the above operations were to make each model's experiment has the same setting to ensure the validity of the comparison results.

It is clear from the Table 4 that the optical flow performed better than the RGB image for all methods. This demonstrated that the results focused more on the tremor changes in the video, which the optical flow could represent better than the RGB information. In addition, all methods' results were significantly improved after EVM pre-processing, which demonstrated the effectiveness of EVM in PD tremor enhancement. In the normal case, X3D performed best with RGB images. While in all other cases, our method achieved the best results. Furthermore, we carried out experiments for two streams (Optical flow + RGB) and found that the results were not better than with optical flow alone, but the calculations increased. Overall, the results demonstrated that our method (Temporal Difference+GSM+ResNet-50 in EVM pre-processing) exhibited a higher accuracy than current state-of-the-art methods in the assessment of PD tremor videos.

Table 4 shows that the accuracy in the EVM pre-processing was much better than in the normal case. To demonstrate the effect of EVM pre-processing on the prediction of video scores, we compared the confusion matrix obtained from the normal case with the EVM pre-processing. The two confusion matrices in Fig. 7 show the prediction results for each score in the normal case and with EVM pre-processing. Comparing the two confusion matrices, the accuracy of all scores improved after EVM pre-processing, which was particularly evident for the scores of 0 and non-0. This suggests that EVM improved the model's sensitivity to the tremors in the videos.

To compare the performance of TSM and GSM in ResNet-50, we visualized their feature maps that were output from one network layer. Fig. 8 shows the feature maps from TSM and GSM in ResNet-50 in the normal case and with EVM pre-processing. These results were obtained from the hand rest tremor experiment. These feature maps were the results of temporal difference from the optical flow, and the temporal changes were due to the tremors in the video. Therefore, the temporal changes in these feature maps were primarily focused on the edge of the hand, which was consistent with the tremor performance in the videos. In comparing the feature maps in normal cases and with EVM pre-processing, it was evident that the temporal changes were significantly enhanced with EVM pre-processing. Furthermore, when

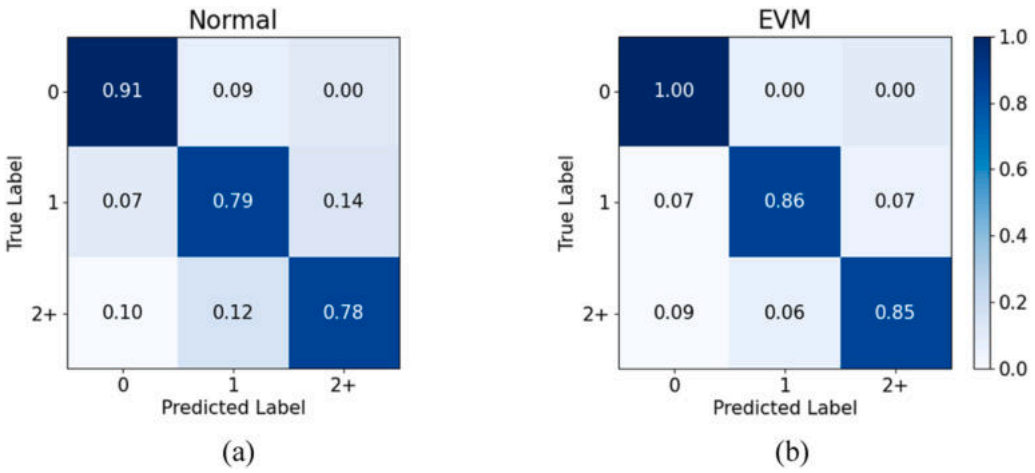

Fig. 7. The comparison of confusion matrix for estimation of the hand rest tremor scores in different cases: (a) confusion matrix in the normal case, (b) confusion matrix in EVM pre-processing case. The accuracy of all scores improved after EVM pre-processing.

comparing TSM and GSM, our proposed GSM makes the temporal changes to be more accurately focused on the edges of the hands, which further illustrated the effectiveness of EVM and GSM for extracting tremor features in the videos.

We presented the results obtained from the three experiments with rest tremors using our method. Table 5 shows the accuracy of multiple classification and binary classification of the different body part. Our method achieved excellent accuracy in predicting the scores of rest tremor videos. The accuracy of binary classification was clearly improved compared to the accuracy of the multiple classifications (scores of 0, 1, 2+). The hand showed the highest accuracy among the three experiments, this is considering that hand tremors are more pronounced than jaw tremors, and the number of videos with different scores for hand tremors was more uniformly distributed. The dichotomous results from these experiments revealed that our method exhibited considerable sensitivity to rest tremor videos for the different body parts. In contrast, it had decreased ability to quantify rest tremor videos. However, the results were in an acceptable range.

Table 6 and Fig. 9 demonstrate the metrics results of our method for the three rest tremor experiments. Our method achieved a macroscopic mean Pre of 0.92, a mean recall of 90%, and an F_1 score of 0.91 for the hand rest tremors. A macroscopic mean Pre of 0.85, an average recall of 85%, and an F_1 score of 0.85 were achieved for the leg rest tremors. A macroscopic mean Pre of 0.91, a mean recall of 86%, and an F_1 score of 0.88 were achieved for the jaw tremors. The average metrics for the leg were lower than for the hands and jaw. The average metrics performed best for the hands. The average metrics for a score of 1 for all parts were typically better than the average metrics of scores 0 and 2+.

Fig. 10 shows the ROC curves with each score for the three experiments. The AUC is the area under the ROC curve, shown as the value of the area in Fig. 10. The micro-average AUC was higher than 0.9 for all three body sites, with a micro-average AUC of 0.93 for the hand, 0.92 for the leg, and 0.96 for the jaw. The jaw achieved the highest micro-average AUC. The micro-average AUC for all parts with scores of 1 was generally lower than the AUC for scores 0 and 2+. The ROC curves for each experiment exhibited satisfactory AUC values.

4.3. Postural tremor score estimation results

In this experiment, we performed score prediction experiments for postural tremors (the 3.15 test in MDS-UPDRS). This part of the experiment was performed only on the videos of the participant's hand, and we compared the results of the Temporal Difference + TSM with GTSN after EVM pre-processing. Table 7 shows that our method outperformed other methods, achieving 84.9% accuracy in the multiple classification and 93.7% accuracy in binary classification. The results indicated that our method was very effective in predicting scores for rest tremors and exhibited high sensitivity in predicting scores for postural tremors.

Fig. 11 shows the confusion matrix for the results of the three scoring classifications used in our method. Although the accuracy of our method remained low for classifications with scores of 1 and 2+, it achieved satisfactory accuracy for the 0 classification, indicating the high sensitivity of our method for postural tremors.

Table 8 and Fig. 12 reveal the evaluation metrics and ROC curves. These average metrics decreased compared to rest tremors but were above 0.85. The micro-average AUC in the ROC curve reached a satisfactory value of 0.93. The reason for the decreased accuracy of postural tremors compared to rest tremors was that postural tremors tended to be accompanied by large movements, which influenced the model assessment. The evaluation metrics of a score of 1 were lower than the evaluation metrics of scores of 0 and 2+, which were the same as the results obtained with rest tremors.

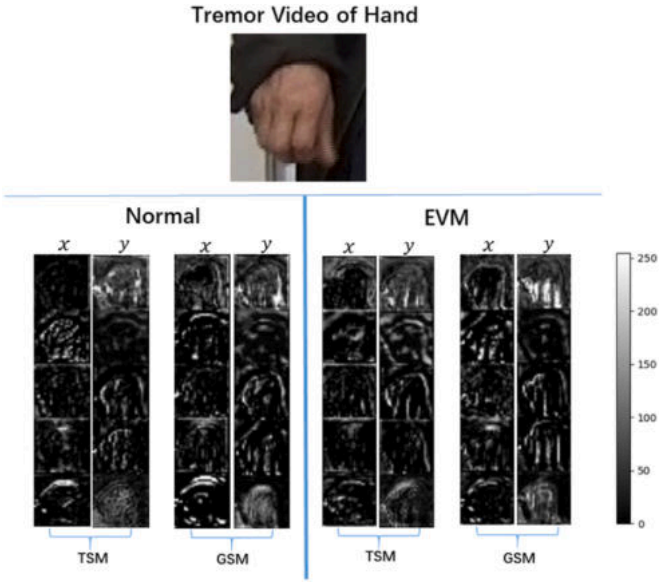

Fig. 8. Optical flow maps visualisation (x and y directions in gray-scales) of our model for hand rest tremor video. The gray value of the pixel is correlated with the tremor at the position. We compare two settings: (1) normal video and EVM pre-processing video; (2) TSM and GSM. In both cases, GSM is more sensitive than TSM to PD tremors in the video.

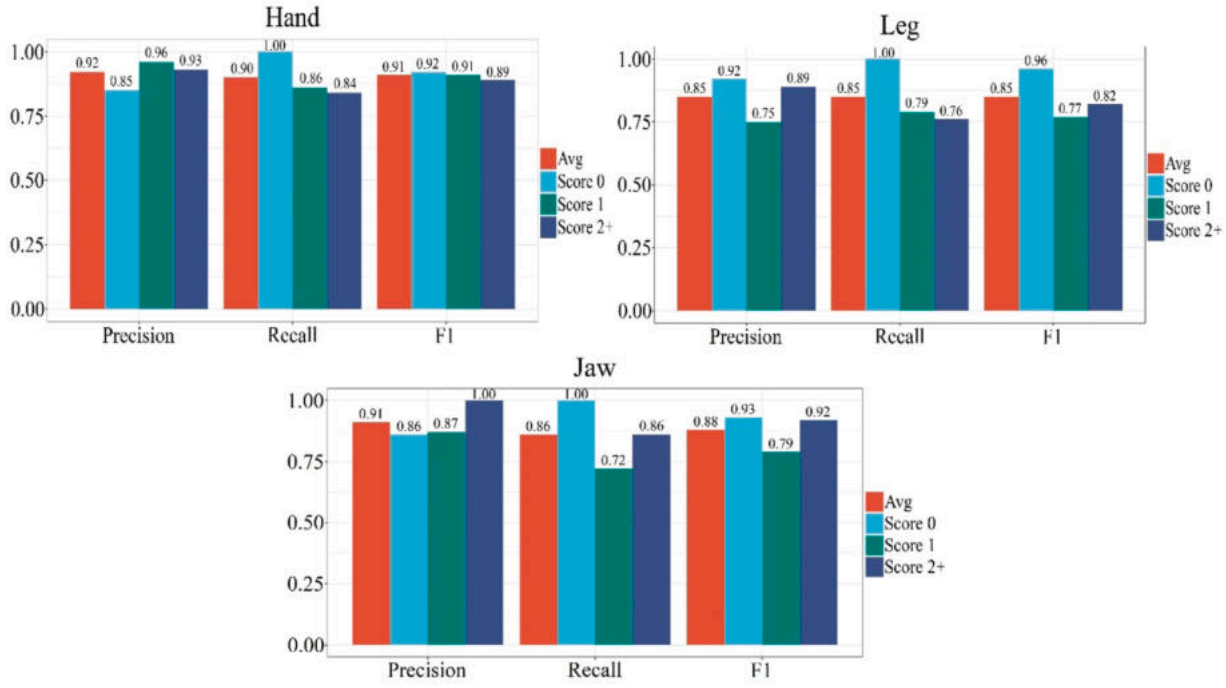

Fig. 9. Assessment performance of our method for per-class of rest tremor score in different body parts. Our method achieves good results in different evaluation metrics, which demonstrates the reliability of our method.

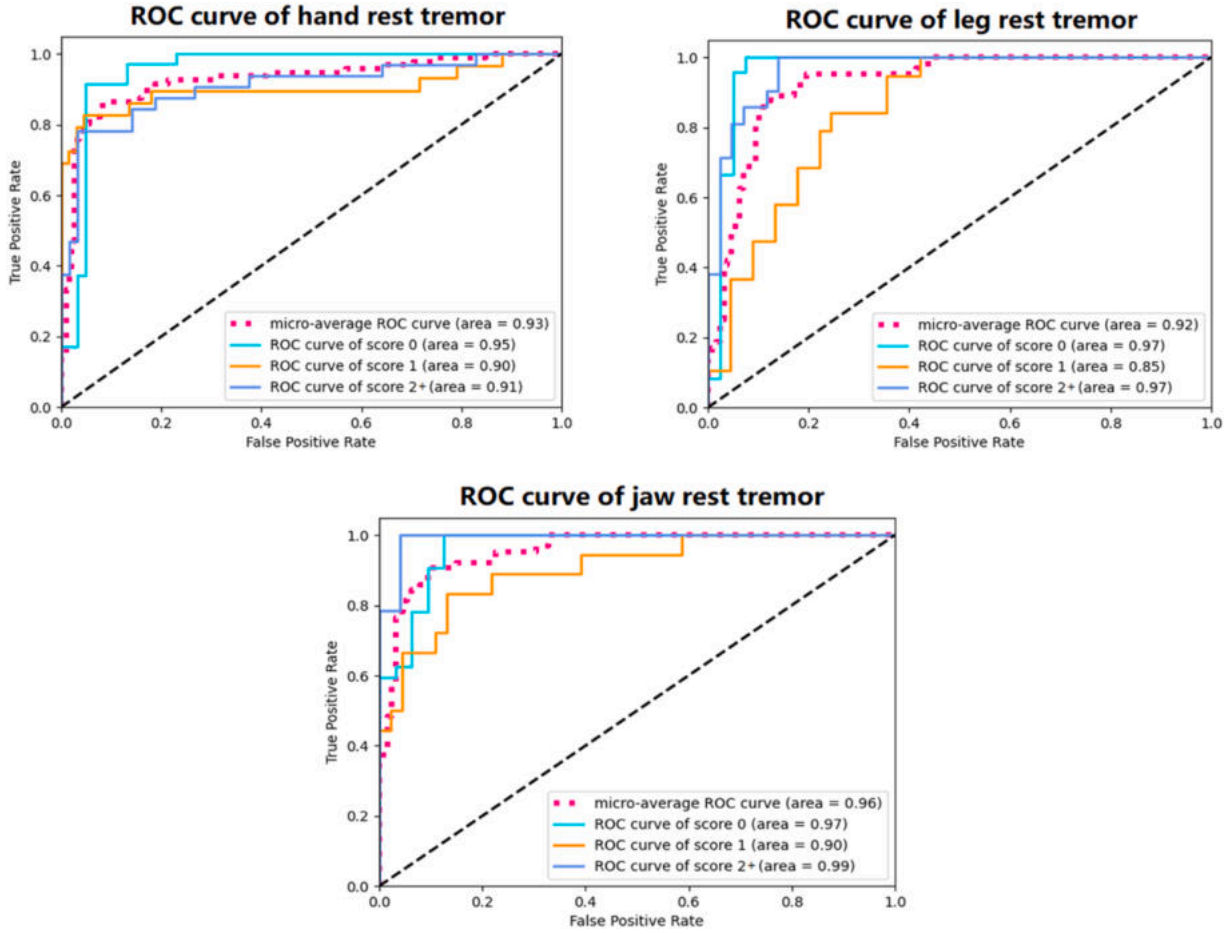

Fig. 10. ROC curves of our method for per-class of rest tremor score in different body parts. The results demonstrate the generalization ability of our method to tremor scores assessment in different body parts.

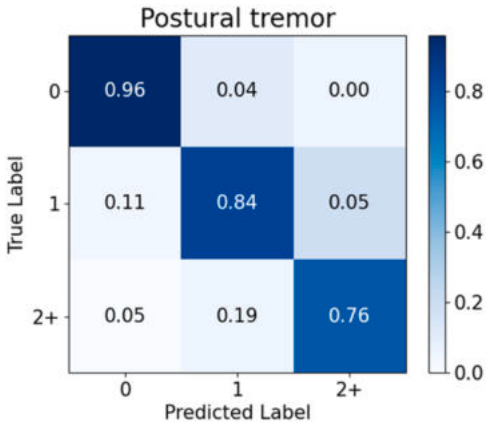

Fig. 11. Confusion matrix of our model for estimation postural tremor scores in EVM pre-processing.

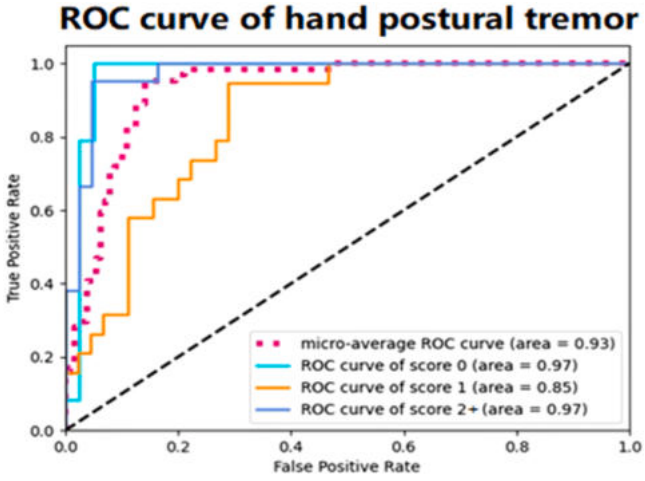

Fig. 12. ROC curves of our method for per-class in postural tremor score.

4.4. More score estimation levels results

To further demonstrate the assessment capability of our model, we extended our assessment level range to (0, 1, 2, and 3+). We combined the scores of 3 and 4 into 3+. As we can see from Tables 1 and 2, both legs and jaw rest tremor lack the videos scoring 3 and 4, so the experiment in this part only contains hand rest tremor and postural tremor. The experiment was based on the previous video data without augmentation. We trained our model under EVM pre-processing without changing the settings.

Through the experiment, we obtained an accuracy of 87.5% for hand rest tremors and 82.2% for hand postural tremors. We presented the detailed results in this section. As can be seen from the experimental results in Table 9 and Fig. 14, the accuracy for non-0 scores (1, 2 and 3+) decrease after extending the assessment level range. However, the result of classification for score 0 still maintains a high accuracy as shown in Fig. 13, which demonstrates that our method is still more advanced than other current methods when it has more assessment levels.

Table 2
Dataset of actual postural tremor participants from 130 participants in this study. The ground-truth for each video is determined by the majority vote of raters. (M: Male; F: Female)

Score	Left Hand Valid/Total	M/F	Right Hand Valid/Total	M/F
0	42/47	21/26	46/50	26/24
1	42/45	26/29	37/38	19/19
2	24/30	14/16	26/29	15/14
3	3/6	4/2	10/11	6/5
4	1/2	2/0	1/2	1/1
All	112/130	67/63	120/130	67/63

Table 3
The actual video dataset of different body parts for the experiment.

Rest Tremor			Postural Tremor
Left Hand/Right Hand	Left Leg/Right Leg	Jaw	Left Hand/Right Hand
256/248	236/236	244	224/240

5. Discussion

In this work, we developed an innovative video-based method to predict scores of PD tremors according to the MDS-UPDRS. Our work provided a novel and reliable way to the assessment of PD tremors, which was suitable for assessing rest and postural tremors. The assessment of the severity of PD tremors is critical to the diagnosis and prognosis of PD, but this is quite a challenging task. Current methods use wearable equipment to extract and analyse PD tremor signals, but such methods could influence PD tremor development due to the weight of the devices (Shawen et al., 2020). In contrast, the model proposed in this study does not cause any interference to the PD tremors, which has a distinct advantage over wearable devices. It ultimately achieved a striking result on a large dataset.

To demonstrate the strengths of our assessment method, we compared our method with the current work on MDS-UPDRS tremor scoring, as seen in Table 10. The current methods using video-based data sources have lower accuracy than the sensor data sources. In addition, most methods are limited to evaluating only a certain type of tremor or a particular body part. Compared to these other methods, our method overcame these limitations. Our video-based method was applied to different body parts with tremors and presented a significant advantage in the accuracy of tremor assessment. Therefore, our method has excellent potential for future clinical assessment and remote monitoring of PD patients. With this method, doctors can easily and quickly measure the severity of the patient's symptoms to formulate better treatment plans.

One critical reason for allowing the MDS-UPDRS score evaluation of PD tremors based on videos is that the video recording is a sampling process of the actual PD tremor, which is consistent with the Nyquist limits. Since the actual scores of PD tremor videos are discrete, predicting the scores of PD tremor videos could be considered a classification process for PD tremor videos. However, the target in most current video classification algorithms tends to be bigger movements, while PD tremors are more detailed and are more subtle movements (Afsar et al., 2015). Therefore, the current video classification algorithms are not very suitable for classifying PD tremor videos. To address this challenge, we designed a temporal difference module to stacks the current optical flow to the result of inter-frame difference so that the model could be more focused on learning information about the tremor. In the field of image and video processing, optical flow represents the motion characteristics of pixels, and RGB image represents the spatial information of images (Reda et al., 2018). However, in the experiments of our study, it was noted that RGB information of image sequences did not improve the sensitivity to tremors based on optical flow. Therefore, we focused on the temporal differences in optical flow to obtain the characteristics of PD tremors in the video.

The scoring criteria of our model for the PD tremor videos were based on trained medical professionals. While we recorded the test video, the evaluators were standing at the same angle and distance as the camera to ensure that the video was consistent with what the evaluators saw. However, the human eye has a distinct advantage in resolution over the common camera we used in our research. Many small PD tremor

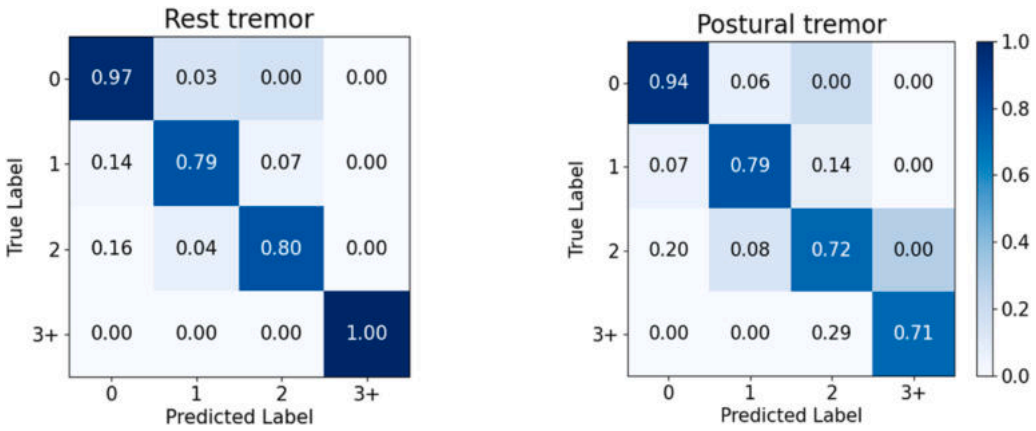

Fig. 13. Confusion matrix of our model for estimation rest and postural tremor scores.

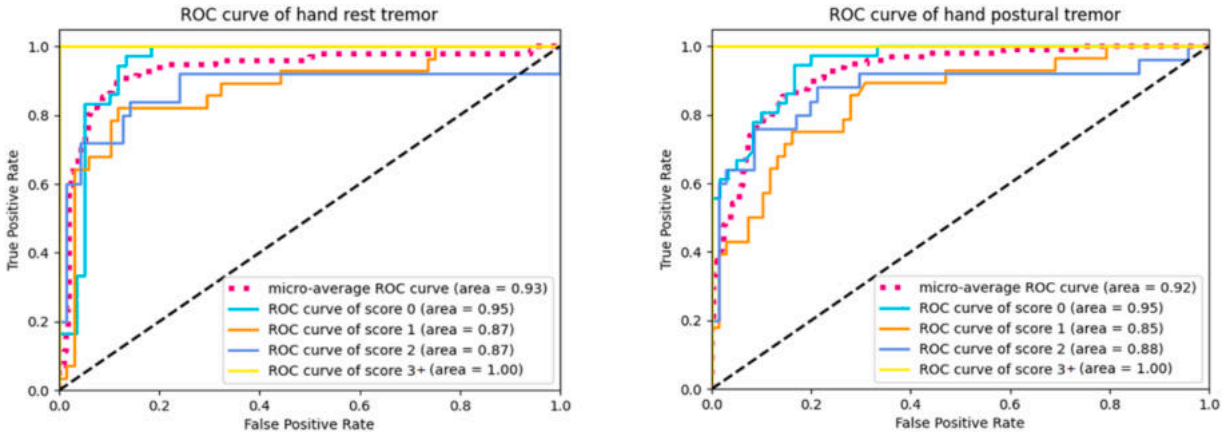

Fig. 14. ROC curves of our method for per-class rest and postural tremor score.

motions could not be shown well in the captured test videos, which could impact model accuracy. EVM amplified the subtle motions in the videos (Abnousi et al., 2019), which enhanced the PD tremors into perceptible size for the model. In the experiment section, we compared the results in normal and EVM pre-processed cases, and the results revealed the effectiveness of EVM in improving the accuracy of the model. The confusion matrix of the results indicated that EVM significantly improved the model's ability to discriminate between scores of 0 and non-0.

In video classification algorithms, TSM makes each temporal segment contain the features of the neighboring temporal segments by simple temporal shifting. In our study, we first considered using this simple but effective module to shift the temporal difference features of tremor videos. However, TSM only allowed the features of the current segment to include the features of the adjacent segments, while the tremor test is a global process. Therefore, we argue that both the features of the adjacent and the no-adjacent segments are correlated to the current segment. GSM extends the scope of the temporal shift to global features so that the features of each segment could include the features of all segments. Based on this experiment, we discovered that GSM exhibited a significant advantage over TSM in predicting PD tremor video scores.

We presented numerous excellent video classification works in Table 4, they have different performances in the classification of PD tremor videos. SlowFast proposed a novel model with slow and fast channels (Feichtenhofer et al., 2019), which is highly sensitive to big movements, as well as fast and slow movements with different frequencies. But compared with other movements, the frequency of PD tremors is more stable, with no obvious change in frequency. Therefore, this method might not be suitable for PD tremor video classification. However, it might perform well in the classification of PD gait videos. I3D performs better in two-stream (Carreira and Zisserman, 2017), but the model has a large amount of parameters that makes it difficult to achieve edge computing. X3D also performs well in PD tremor assessment with less parameters, but its current work only focuses on RGB stream of video (Feichtenhofer, 2020). We believe it might have good potential in PD severity assessment.

Furthermore, the model's calculation also was an important factor in PD severity research. For the frame sampling of PD tremor videos, a higher fps might contain more subtle features of tremors, which could slightly improve the results. However, this also would increase the input data and model calculations. Therefore, it was necessary to sample PD tremor videos. As seen in Table 4, the Optical flow+RGB results did not significantly outperform the Optical flow results. The reasons for this observation might be that the feature fusion of tremors in two stream networks did not perform well, or the RGB stream might include some noise in addition to the tremor features. In addition, although some methods receive a small enhancement in the Optical flow+RGB cases, a large increase in the number of calculations and parameters to obtain only a minor improvement in the results was not desirable.

From the ROC curve of the experimental results, we found that our model has a stronger generalization ability for the most severe score. We believe this is due to the fact that higher scores mean the tremors are more pronounced and easier to distinguish, which is consistent with the tremor scoring rules in the MDS-UPDRS. In the experiments with more score estimation levels, it can be found surprising accuracy for the most severe score of 3+. One of the reasons is our method has a stronger generalization ability for more severe videos. On the other hand, it is due

Table 4
Comparison of different methods' accuracy. RGB and optical flow comparison in normal and EVM pre-processed cases.

Pre-processing	Method	Backbone	ACC(%)		
			RGB	Optical flow	RGB+Optical flow
Normal	Temporal Difference+I3D	ResNet-50	73.9	76.0	76.9
	Temporal Difference+SlowFast	ResNet-50	72.9	——	——
	Temporal Difference+X3D	X3D-XL	**83.5**	——	——
	Temporal Difference+TSN	ResNet-50	77.3	78.9	78.4
	Temporal Difference+TSM	ResNet-50	79.8	80.2	80.5
	Temporal Difference+GSM **(Ours)**	ResNet-50	81.0	**83.3**	**83.2**
EVM	Temporal Difference+I3D	ResNet-50	77.1	79.9	80.1
	Temporal Difference+SlowFast	ResNet-50	76.0	——	——
	Temporal Difference+X3D	X3D-XL	87.5	——	——
	Temporal Difference+TSN	ResNet-50	79.6	81.4	81.2
	Temporal Difference+TSM	ResNet-50	80.5	83.3	83.0
	Temporal Difference+GSM **(Ours)**	ResNet-50	**88.7**	**90.6**	**91.2**

Table 5
The accuracy of our method for different body parts rest tremor. Comparison of multiple classification ((scores of 0, 1, 2+)) and binary classification.

	Multiple-ACC(%)	Binary-ACC(%)
Hand	90.6	94.8
Leg	85.9	96.9
Jaw	89.0	95.3

Table 9
Comparison of metrics results for hand rest tremor and postural tremor. The results include each score and the average.

Score	Rest tremor			Postural tremor		
	Pre	Rec	F_1	Pre	Rec	F_1
0	0.81	0.97	0.89	0.83	0.94	0.88
1	0.92	0.79	0.85	0.85	0.79	0.81
2	0.91	0.80	0.85	0.75	0.72	0.73
3+	1.00	1.00	1.00	1.00	0.71	0.83
Avg	0.91	0.89	0.90	0.86	0.79	0.82

Table 6
Comparison of metrics results for different body parts. The results include each score and the average.

	Pre				Recall				F_1			
	0	1	2+	Avg	0	1	2+	Avg	0	1	2+	Avg
Hand	0.85	0.96	0.93	0.92	1.00	0.86	0.84	0.90	0.92	0.91	0.89	0.91
Leg	0.92	0.75	0.89	0.85	1.00	0.79	0.76	0.85	0.96	0.77	0.82	0.85
Jaw	0.86	0.87	1.00	0.91	1.00	0.72	0.86	0.86	0.93	0.79	0.92	0.88

Table 7
Accuracy of our method for postural tremor. Comparison of multiple classification ((scores of 0, 1, 2+)) and binary classification.

Method	Multiple-ACC	Binary-ACC
EVM+Temporal Difference+TSM	80.6	89.1
EVM+Ours	**84.9**	**93.7**

Table 8
Comparison of metrics results for postural tremor. The results include each score and the average.

Score	Pre	Rec	F_1
0	0.88	0.96	0.92
1	0.76	0.84	0.80
2+	0.94	0.76	0.84
Avg	0.86	0.85	0.85

to the small number of videos with score of 3+. Even though we use data enhancement and cross-validation methods, it is still possible more positive samples in very few videos with score of 3+, which may lead to the very ideal results.

Our study still has some room for improvement. The PD tremor test in MDS-UPDRS was aimed at different body parts since each part has different tremor patterns. We trained the tremor video of each body part independently, which might have increased the workload associated with the model training. We comprehensively compared the performance of our method and the wearable method in the PD tremor assessment. Vision-based methods are very popular since they are more affordable and allow unconstrained limb movements. In the PD videos, limbs are likely to be occluded during interaction, while wearable devices can get the tremor signal directly. However, wearable devices can negatively affect the results due to their weight. Therefore, both of the two methods could find their usage under different circumstances in PD tremor assessment and were still in constant development. For future research, our method has significant advantages in monitoring patients' conditions, while the wearable method is more suitable for analysing the tremor electrophysiological mechanism. Furthermore, participants' movements needed to be restricted from following the rules of the MDS-UPDRS to obtain accurate videos. Therefore, future application of this work to the daily monitoring of PD patients might require additional consideration of how to obtain ideal videos of routine behaviors.

A common challenge in PD severity assessment studies is the lack of data with mild and moderate severity. We also faced this issue in our work, which can be seen in Tables 1 and 2. Although we performed data augmentation on these videos, the datasets were still unevenly distributed. In the experimental results, the accuracy of binary classification (0 and non-0) was excellent, while the accuracy of multiple classification was less well, which was consistent with the expectations. From the evaluators' perspective, this was due to the scoring criteria of the model that were based on the empirical judgment of the evaluators. The score of non-0 was a subjective quantitative process, so it was difficult to distinguish. Although we confirmed the final video scores by majority

Table 10
Comparison of other methods in MDS-UPDRS scoring of tremors tests.

Methods	Data source	Participants	Tremor	Object	Score	ACC(%)
Rigas et al. (2012)	Sensor	23	Rest, Postural	Hand, Leg	(0,1,2,3)	87.0
Kim et al. (2018)	Sensor	92	Rest	Hand	(0,1,2,3+)	85.0
Chang et al. (2019)	Videos	106	Rest	Hand	(0,1)	72.2
Yin et al. (2022)	Videos	85	Postural	Hand	(0,1)	70.6
	Videos	85	Postural	Hand	(0,1,2)	60.8±3.5
Ours	**Videos**	**130**	**Rest, Postural**	**Hand, Leg, Jaw**	**(0,1,2+)**	**84.9~90.6**
	Videos	**130**	**Rest, Postural**	**Hand**	**(0,1,2,3+)**	**82.2~87.5**

vote, the description of the tremor severity in the MDS-UPDRS was still somewhat subjective.

6. Conclusion

Herein, we proposed a video-based method to predict the scores of rest and postural tremors in MDS-UPDRS. This research focused on the continuous and subtle PD tremors in the video. Due to the limited capability of common cameras in capturing motion, we adapted EVM to the video pre-processing to magnify the subtle tremors that were difficult to represent in the videos. Current video classification algorithms were not suitable to accurately assess the tremor videos. To solve this issue, we proposed GTSN, a model that focuses more on the micro temporal changes caused by the tremors. Considering that the scoring of PD tremors is a global process, we propose the plug-and-play GSM that allowed the features of the current temporal segment to include features of the global temporal segment, which significantly increased the model's prediction accuracy. The effectiveness of our proposed method (EVM+GTSN) for the score prediction of PD tremor videos was demonstrated through experiments. This work could be used for supplementary assessment of PD tremors to reduce the stress of the healthcare system. In addition, the method could be considered as an approach for the classification of other subtle motion videos in the future.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This study was supported by the United Fujian Provincial Health and Education Project for Tackling the Key Research of China (2019-WJ-03); the Special Funds of the Central Government Guiding Local Science and Technology Development (2020L3008)

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.media.2023.102754.

References

Abnousi, F., Kang, G., Giacomini, J., Yeung, A., Zarafshar, S., Vesom, N., Ashley, E., Harrington, R., Yong, C., 2019. A novel noninvasive method for remote heart failure monitoring: the EuleriAn video Magnification apPLications In heart Failure studY (AMPLIFY). NPJ Digit. Med. 2, 1–6.

Afsar, P., Cortez, P., Santos, H., 2015. Automatic visual detection of human behavior: a review from 2000 to 2014. Expert Syst. Appl. 42, 6935–6956.

Ali, M.R., Hernandez, J., Dorsey, E.R., Hoque, E., McDuff, D., 2020. Spatio-temporal attention and magnification for classification of Parkinson's disease from videos collected via the internet. In: Proceedings of the 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG). Buenos Aires, ARGENTINA, pp. 207–214.

Amoroso, N., La Rocca, M., Monaco, A., Bellotti, R., Tangaro, S., 2018. Complex networks reveal early MRI markers of Parkinson's disease. Med. Image Anal. 48, 12–24.

Baumann, C.R., 2012. Epidemiology, diagnosis and differential diagnosis in Parkinson's disease tremor. Parkinsonism Relat. Disord. 18 (Suppl 1), S90–S92.

Bhatia, K.P., Bain, P., Bajaj, N., Elble, R.J., Hallett, M., Louis, E.D., Raethjen, J., Stamelou, M., Testa, C.M., Deuschl, G., Int Parkinson Movement, D., 2018. Consensus statement on the classification of tremors. From the task force on tremor of the international parkinson and movement disorder society. Mov. Disord. 33, 75–87.

Bhattacharjee, S., Sambamoorthi, U., 2013. Co-occurring chronic conditions and healthcare expenditures associated with Parkinson's disease: a propensity score matched analysis. Parkinsonism Relat. Disord. 19, 746–750.

Bi, X.-A., Hu, X., Xie, Y., Wu, H., 2021. A novel CERNNE approach for predicting Parkinson's disease-associated genes and brain regions based on multimodal imaging genetics data. Med. Image Anal. 67, 101830.

Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., Sheikh, Y., 2021. OpenPose: realtime multi-person 2D pose estimation using part affinity fields. IEEE Trans. Pattern Anal. Mach. Intell. 43, 172–186.

Carreira, J., Zisserman, A., 2017. Quo vadis, action recognition? A new model and the kinetics dataset. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6299–6308.

Chang, C.M., Huang, Y.L., Chen, J.C., Lee, C.C., 2019. Improving automatic tremor and movement motor disorder severity assessment for Parkinson's disease with deep joint training. In: Proceedings of the 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 3408–3411.

Dai, H., Cai, G., Lin, Z., Wang, Z., Ye, Q., 2021. Validation of inertial sensing-based wearable device for tremor and bradykinesia quantification. IEEE J. Biomed. Health Inform. 25, 997–1005.

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, 1837–1845.

Dentamaro, V., Impedovo, D., Pirlo, G., 2020. Gait analysis for early neurodegenerative diseases classification through the kinematic theory of rapid human movements. IEEE Access 8, 193966–193980.

Dong, L., Zheng, Y.M., Luo, X.G., He, Z.Y., 2021. High inflammatory tendency induced by malignant stimulation through imbalance of CD28 and CTLA-4/PD- I contributes to dopamine neuron injury. J. Inflamm. Res. 14, 2471–2482.

Duval, C., Beuter, A., 1998. Fluctuations in tremor at rest and eye movements during ocular fixation in subjects with Parkinson's disease. Parkinsonism Relat. Disord. 4, 91–97.

Feichtenhofer, C., 2020. X3D: expanding architectures for efficient video recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 203–213.

Feichtenhofer, C., Fan, H., Malik, J., He, K., 2019. Slowfast networks for video recognition. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 6202–6211.

Fino, P.C., Mancini, M., 2020. Phase-dependent effects of closed-loop tactile feedback on gait stability in Parkinson's disease. IEEE Trans. Neural Syst. Rehabil. Eng. 28, 1636–1641.

Florin, E., Reck, C., Burghaus, L., Lehrke, R., Gross, J., Sturm, V., Fink, G.R., Timmermann, L., 2008. Ten Hertz thalamus stimulation increases tremor activity in the subthalamic nucleus in a patient with Parkinson's disease. Clin. Neurophysiol. 119, 2098–2103.

Goetz, C.G., Stebbins, G.T., Tilley, B.C., 2012. Calibration of unified Parkinson's disease rating scale scores to movement disorder society-unified Parkinson's disease rating scale scores. Mov. Disord. 27, 1239–1242.

Guo, R., Shao, X., Zhang, C., Qian, X., 2020. Sparse adaptive graph convolutional network for leg agility assessment in Parkinson's disease. IEEE Trans. Neural Syst. Rehabil. Eng. 28, 2837–2848.

Hughes, J.A., Houghten, S., Brown, J.A., 2020. Models of Parkinson's disease patient gait. IEEE J. Biomed. Health Inform. 24, 3103–3110.

Kim, H.B., Lee, W.W., Kim, A., Lee, H.J., Park, H.Y., Jeon, H.S., Kim, S.K., Jeon, B., Park, K.S., 2018. Wrist sensor-based tremor severity quantification in Parkinson's disease using convolutional neural network. Comput. Biol. Med. 95, 140–146.

Kuosmanen, E., Wolling, F., Vega, J., Kan, V., Nishiyama, Y., Harper, S., Van Laerhoven, K., Hosio, S., Ferreira, D., 2020. Smartphone-based monitoring of

Parkinson disease: quasi-experimental study to quantify hand tremor severity and medication effectiveness. JMIR mHealth uHealth 8, e21543.
Lee, M., Lee, S., Son, S., Park, G., Kwak, N., 2018. Motion feature network: fixed motion filter for action recognition. In: Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich, GERMANY, pp. 392–408.
Legaria-Santiago, V.K., Sanchez-Fernandez, L.P., Sanchez-Perez, L.A., Garza-Rodriguez, A., 2022. Computer models evaluating hand tremors in Parkinson's disease patients. Comput. Biol. Med. 140, 105059.
Li, H., Shao, X., Zhang, C., Qian, X., 2021. Automated assessment of Parkinsonian finger-tapping tests through a vision-based fine-grained classification model. Neurocomputing 441, 260–271.
Li, M.H., Mestre, T.A., Fox, S.H., Taati, B., 2018a. Automated assessment of levodopa-induced dyskinesia: evaluating the responsiveness of video-based features. Parkinsonism Relat. Disord. 53, 42–45.
Li, M.H., Mestre, T.A., Fox, S.H., Taati, B., 2018b. Vision-based assessment of Parkinsonism and levodopa-induced dyskinesia with pose estimation. J. NeuroEng. Rehabil. 15, 1–13.
Li, X., Xing, Y., Martin-Bastida, A., Piccini, P., Auer, D.P., 2018c. Patterns of grey matter loss associated with motor subscores in early Parkinson's disease. Neuroimage-Clin. 17, 498–504.
Liddle, J., Ireland, D., McBride, S.J., Brauer, S.G., Hall, L.M., Ding, H., Karunanithi, M., Hodges, P.W., Theodoros, D., Silburn, P.A., Chenery, H.J., 2014. Measuring the lifespace of people with Parkinson's disease using smartphones: proof of principle. JMIR mHealth uHealth 2, e2799.
Lin, J., Gan, C., Wang, K., Han, S., 2020. TSM: temporal shift module for efficient and scalable video understanding on edge devices. IEEE Trans. Pattern Anal. Mach. Intell. 44, 2760–2774.
Liu, C., Li, Z., Chang, F., Li, S., Xie, J., 2021a. Temporal shift and spatial attention-based two-stream network for traffic risk assessment. IEEE Trans. Intell. Transp. Syst. 23, 12518–12530.
Liu, G., Zhang, Q., Cao, Y., Tian, G., Ji, Z., 2021b. Online human action recognition with spatial and temporal skeleton features using a distributed camera network. Int. J. Intell. Syst. 36, 7389–7411.
Liu, J., Akhtar, N., Mian, A., 2018. Viewpoint Invariant Action recognition using RGB-D videos. IEEE Access 6, 70061–70071.
Liu, Y., Chen, J., Hu, C., Ma, Y., Ge, D., Miao, S., Xue, Y., Li, L., 2019. Vision-based method for automatic quantification of Parkinsonian bradykinesia. IEEE Trans. Neural Syst. Rehabil. Eng. 27, 1952–1961.
Lu, M., Poston, K., Pfefferbaum, A., Sullivan, E.V., Adeli, E., 2020. Vision-based estimation of MDS-UPDRS gait scores for assessing Parkinson's disease motor severity. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 637–647.
Lu, M., Zhao, Q., Poston, K.L., Sullivan, E.V., Pfefferbaum, A., Shahid, M., Katz, M., Kouhsari, L.M., Schulman, K., Milstein, A., Niebles, J.C., Henderson, V.W., Li, F.F., Pohl, K.M., Adeli, E., 2021. Quantifying Parkinson's disease motor severity under uncertainty using MDS-UPDRS videos. Med. Image Anal. 73, 102179.
Luca, M., Giovanni, R., Carlo, C., 2018. Diagnostic Criteria for Parkinson's disease: from james Parkinson to the concept of prodromal disease. Front. Neurol. 9, 156.
Lukhanina, E.P., Kapoustina, M.T., Karaban, I.N., 2000. A quantitative surface electromyogram analysis for diagnosis and therapy control in Parkinson's disease. Parkinsonism Relat. Disord. 6, 77–86.
Marxreiter, F., Buttler, U., Gassner, H., Gandor, F., Gladow, T., Eskofier, B., Winkler, J., Ebersbach, G., Klucken, J., 2020. The use of digital technology and media in German Parkinsons disease patients. J. Parkinsons Dis. 10, 717–727.
Massano, J., Bhatia, K.P., 2012. Clinical approach to Parkinson's disease: features, diagnosis, and principles of management. Cold Spring Harb. Perspect. Med. 2, a008870.
Mei, J., Desrosiers, C., Frasnelli, J., 2021. Machine learning for the diagnosis of Parkinson's disease: a review of literature. Front. Aging Neurosci. 13, 633752.
Monje, M.H.G., Foffani, G., Obeso, J., Sanchez-Ferro, A., Yamush, M.L., 2019. New Sensor and Wearable Technologies to Aid in the Diagnosis and Treatment Monitoring of Parkinson's Disease, 21. Annual Review of Biomedical Engineering, pp. 111–143.
Ong, M.-S., Magrabi, F., Coiera, E., 2012. Automated identification of extreme-risk events in clinical incident reports. J. Am. Med. Inform. Assoc. 19, E110–E118.
Pintea, S.L., Zheng, J., Li, X., Bank, P.J.M., van Hilten, J.J., van Gemert, J.C., 2018. Hand-tremor frequency estimation in videos. In: Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich, GERMANY, pp. 213–228.
Politis, M., Wu, K., Molloy, S., Bain, P.G., Chaudhuri, K.R., Piccini, P., 2010. Parkinson's disease symptoms: the patient's perspective. Mov. Disord. 25, 1646–1651.
Porritt, M.J., Kingsbury, A.E., Hughes, A.J., Howells, D.W., 2006. Striatal dopaminergic neurons are lost with Parkinson's disease progression. Mov. Disord. 21, 2208–2211.
Reda, F.A., Liu, G., Shih, K.J., Kirby, R., Barker, J., Tarjan, D., Tao, A., Catanzaro, B., 2018. SDC-net: video prediction using spatially-displaced convolution. In: Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich, GERMANY, pp. 747–763.
Rigas, G., Tzallas, A.T., Tsipouras, M.G., Bougia, P., Tripoliti, E.E., Baga, D., Fotiadis, D. I., Tsouli, S.G., Konitsiotis, S., 2012. Assessment of tremor activity in the Parkinson's disease using a set of wearable sensors. IEEE Trans. Inf. Technol. Biomed. 16, 478–487.
Rupprechter, S., Morinan, G., Peng, Y., Foltynie, T., Sibley, K., Weil, R.S., Leyland, L.-A., Baig, F., Morgante, F., Gilron, R.e., Wilt, R., Starr, P., Hauser, R.A., O'Keeffe, J., 2021. A clinically interpretable computer-vision based method for quantifying gait in Parkinson's disease. Sensors 21, 5437.
Sabo, A., Mehdizadeh, S., Ng, K.D., Iaboni, A., Taati, B., 2020. Assessment of Parkinsonian gait in older adults with dementia via human pose tracking in video data. J. NeuroEng. Rehabil. 17, 1–10.
Shawen, N., O'Brien, M.K., Venkatesan, S., Lonini, L., Simuni, T., Hamilton, J.L., Ghaffari, R., Rogers, J.A., Jayaraman, A., 2020. Role of data measurement characteristics in the accurate detection of Parkinson's disease symptoms using wearable sensors. J. Neuroeng. Rehabil. 17, 1–14.
Silva de Lima, A.L., Evers, L.J.W., Hahn, T., Bataille, L., Hamilton, J.L., Little, M.A., Okuma, Y., Bloem, B.R., Faber, M.J., 2017. Freezing of gait and fall detection in Parkinson's disease using wearable sensors: a systematic review. J. Neurol. 264, 1642–1654.
Stebbins, G.T., Goetz, C.G., Burn, D.J., Jankovic, J., Khoo, T.K., Tilley, B.C., 2013. How to identify tremor dominant and postural instability/gait difficulty groups with the movement disorder society unified Parkinson's disease rating scale: comparison with the unified Parkinson's disease rating scale. Mov. Disord. 28, 668–670.
Ureten, K., Maras, H.H., 2022. Automated classification of rheumatoid arthritis, osteoarthritis, and normal hand radiographs with deep learning methods. J. Digit. Imaging 35, 193–199.
Vig, E., Dorr, M., Cox, D.D., 2012. Saliency-based selection of sparse descriptors for action recognition. In: Proceedings of the 19th IEEE International Conference on Image Processing (ICIP). Lake Buena Vista, FL, pp. 1405–1408.
Vignoud, G., Desjardins, C., Salardaine, Q., Mongin, M., Garcin, B., Venance, L., Degos, B., 2022. Video-based automated analysis of MDS-UPDRS III parameters in Parkinson disease. Biorxiv.
Voillemin, T., Wannous, H., Vandeborre, J.-P., 2021. 2D deep video capsule network with temporal shift for action recognition. In: Proceedings of the 25th International Conference on Pattern Recognition (ICPR). Electr Network, pp. 3513–3519.
Wang, L., Huynh, D.Q., Koniusz, P., 2020. A comparative review of recent kinect-based action recognition algorithms. IEEE Trans. Image Process. 29, 15–28.
Wang, L., Xiong, Y., Wang, Z., Qiao, Y., Lin, D., Tang, X., Van Gool, L., 2019. Temporal segment networks for action recognition in videos. IEEE Trans. Pattern Anal. Mach. Intell. 41, 2740–2755.
Wang, Q., Liu, W., Chen, X., Wang, X., Chen, G., Zhu, X., 2021a. Quantification of scar collagen texture and prediction of scar development via second harmonic generation images and a generative adversarial network. Biomed. Opt. Express 12, 5305–5319.
Wang, Q., Liu, W., Wang, X., Chen, X., Chen, G., Wu, Q., 2022. A spatial-temporal graph model for pronunciation feature prediction of Chinese poetry. IEEE Trans. Neural Netw. Learn. Syst. 1–15.
Wang, Q., Zhang, Y., Chen, G., Chen, Z., Hee, H.I., 2021b. Assessment of heart rate and respiratory rate for perioperative infants based on ELC model. IEEE Sensors J. 21, 13685–13694.
Wang, X., Garg, S., Tran, S.N., Bai, Q., Alty, J., 2021c. Hand tremor detection in videos with cluttered background using neural network based approaches. Health Inf. Sci. Syst. 9, 1–14.
Wu, H.-Y., Rubinstein, M., Shih, E., Guttag, J., Durand, F., Freeman, W., 2012. Eulerian video magnification for revealing subtle changes in the world. ACM Trans. Graph. 31, 1–8.
Yin, Z., Geraedts, V.J., Wang, Z., Contarino, M.F., Dibeklioglu, H., van Gemert, J., 2022. Assessment of Parkinson's disease severity from videos using deep architectures. IEEE J. Biomed. Health Inform. 26, 1164–1176.
Zach, H., Dirkx, M., Bloem, B.R., Helmich, R.C., 2015. The Clinical Evaluation of Parkinson's Tremor. J. Parkinsons Dis. 5, 471–474.
Zhang, H., Ho, E.S.L., Zhang, X., Shum, H.P.H., 2022. Pose-based tremor classification for Parkinson's disease diagnosis from video, p. arXiv:2207.06828.
Zhao, L., Xu, J., Gong, C., Yang, J., Zuo, W., Gao, X., 2021. Learning to acquire the quality of human pose estimation. IEEE Trans. Circuits Syst. Video Technol. 31, 1555–1568.
Zhou, Y., Yu, H., Wang, S., 2017. Feature sampling strategies for action recognition. In: Proceedings of the 24th IEEE International Conference on Image Processing (ICIP). Beijing, PEOPLES R CHINA, pp. 3968–3972.
Zitser, J., Peretz, C., David, A.B., Shabtai, H., Ezra, A., Kestenbaum, M., Brozgol, M., Rosenberg, A., Herman, T., Balash, Y., Gadoth, A., Thaler, A., Stebbins, G.T., Goetz, C.G., Tilley, B.C., Luo, S.T., Liu, Y., Giladi, N., Gurevich, T., 2017. Validation of the hebrew version of the movement disorder society-unified Parkinson's disease rating scale. Parkinsonism Relat. Disord. 45, 7–12.