



Deposited via The University of Leeds.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/242443/>

Version: Accepted Version

Proceedings Paper:

Liu, W., Quinn, T., Williams, S. et al. (Accepted: 2026) From Optical Flow to Motion Magnitude: An Efficient Framework for Video-Based Hand Tremor Assessment. In: Proceedings of CBMS 2026. The 39th IEEE International Symposium on Computer-Based Medical Systems IEEE CBMS 2026, 03-05 Jun 2026, Limassol, Cyprus. IEEE. (In Press)

This is an author produced version of a conference paper accepted for publication in the Proceedings of CMBS 2026, made available via the University of Leeds Research Outputs Policy under the terms of the Creative Commons Attribution License (CC-BY), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

From Optical Flow to Motion Magnitude: An Efficient Framework for Video-Based Hand Tremor Assessment

Weiping Liu^{1,4}, Thomas Quinn¹, Stefan Williams², Jane Alty³, Samuel Relton¹, David Wong¹

1. *Leeds Institute of Health Sciences, University of Leeds, Leeds, United Kingdom*

2. *Department of Neurology, Leeds Teaching Hospitals Trust, Leeds, United Kingdom*

3. *Wicking Dementia Research and Education Centre, University of Tasmania, Hobart, Australia*

4. *Fujian Key Laboratory of Pattern Recognition and Image Understanding, Xiamen University of Technology, Xiamen, China*

Abstract—Tremor is a characteristic sign of multiple neurological conditions and plays a critical role in assessment of disease severity and prognosis. Recent computer vision-based approaches using optical flow have demonstrated promising performance for video-based tremor evaluation. However, conventional optical flow representations encode directional components separately, which may introduce redundant computations and directional noise. In this study, we propose a motion magnitude-based framework to replace the conventional two-channel optical flow representation for modeling tremor dynamics in video sequences. The motion magnitude, derived from optical flow, provides a direction-invariant representation that better aligns with tremor amplitude characteristics while reducing computational complexity. To effectively exploit this representation, we design a dedicated feature extraction architecture tailored for magnitude-based inputs.

Models were trained and tested on a cohort of 65 patients with pathological resting tremor, from which 130 videos were obtained. Each video was graded by a neurology consultant on a score of 0-4 using the relevant part of the Movement Disorder Society Unified Parkinson’s Disease Rating Scale (MDS-UPDRS). Our experiments demonstrate that the proposed method outperformed previous optical flow-based approaches. The framework achieved an overall accuracy of (71.76%) and a within-one-grade accuracy of (98.35%) compared to the human rater. These results indicate that the proposed motion magnitude representation provides an efficient and effective solution for automated tremor severity evaluation. The source code is publicly available at https://github.com/lwp-1995/Motion_Magnitude_PD_Tremor.

Index Terms—Tremor, MDS-UPDRS, Computer Vision, Optical Flow

I. INTRODUCTION

Rhythmic involuntary movements of the hand, known as tremor, are a sign of multiple conditions including Parkinson’s disease (PD), essential tremor and multiple sclerosis. Tremor is clinically assessed via human expert judgment, by considering its frequency and amplitude. Such assessment is important during disease diagnosis, for disease monitoring, and to determine the effectiveness of medication. In research settings,

tremor is rated on numerical scales. Two popular scales that include hand tremor, the Movement Disorder Society-Unified Parkinson’s Disease Rating Scale (MDS-UPDRS) and The Essential Tremor Rating Scale (TETRAS), assess tremor severity via amplitude on a 0-4 scale in which 4 represents the most severe tremor [1], [2]. While manual judgement using these scales is commonplace, previous work has shown that one weakness of human assessment is inconsistency between raters [3].

Computer vision methods have become increasingly popular as an approach to objectively assess motor signs, including tremor, from video recordings [4], [5]. The majority of existing methods record a video of the trembling hand, and then convert the video into a set of key markers using pose estimation pre-trained models such as Mediapipe [6]. While this approach has shown promise, pre-trained pose estimation models can produce poor results when the hand is partly occluded [7], and when movement between frames is small such that pose-estimation error is of the similar magnitude to tremor movement. There is also initial evidence suggesting that such models exhibit differing accuracy dependent on skin tone [8].

One alternative is to calculate optical flow directly from the raw video [9]. Optical flow is a computer vision method that estimates local pixel velocity by assuming that the brightness of a point on an object remains constant over subsequent frames. In our previous work, we proposed a video classification-based framework for the accurate quantification and assessment of PD tremor severity [10]. We demonstrated that optical flow-based representations outperform RGB-based methods in this setting, as optical flow effectively captures motion dynamics and temporal variations in video streams. The aim of this study is to improve on previous work by investigating the impact of using motion magnitude from the two directional components of optical flow. Compared with directly using optical flow as input, motion magnitude provides a direction-invariant representation of tremor, which is more consistent with clinical assessment criteria. Meanwhile, it enhances robustness to viewpoint variations and mitigates

This work was supported by the National Institute for Health Research (NIHR) I4I Program (NIHR203399). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

noise introduced by unstable directional flow components.

The main contributions of this work are to:

- 1) Outline a computer vision-based framework for automatic hand tremor assessment from video data. Compared with previous approaches, the proposed method achieves improved performance while reducing computational complexity.
- 2) Refine conventional optical flow-based representations by introducing a motion magnitude formulation. We demonstrate that motion magnitude provides a more effective representation of tremor dynamics than the original two-channel optical flow.
- 3) fully exploit the motion magnitude representation, we design an effective feature extraction module and backbone architecture tailored for magnitude-based inputs, enabling robust tremor severity prediction.

II. RELATED WORK

Most current computer vision work on PD tremor can be grouped into two broad lines: skeleton-based methods that operate on pose keypoints, and RGB- and optical-flow-based methods that exploit raw appearance and dense motion fields to capture subtle shaking [11].

A. Skeleton-Based Methods

Skeleton-based hand tremor assessment typically applies pose estimators, such as OpenPose [12] and MediaPipe Hands [6], to monocular RGB videos to extract joint trajectories. These keypoints are then used to compute kinematic, time-domain, and frequency-domain features for tremor frequency and severity estimation. For Parkinson's Disease (PD) tremor, SPAPNet [13] and SPA-PTA [14] use lightweight pyramidal attention to highlight tremor-related pose dynamics from everyday videos, showing strong results in both binary detection and multi-level severity scoring. To move beyond hand-crafted features, recent approaches model skeleton sequences end-to-end using more expressive graph operators that capture rich spatio-temporal dependencies between joints. Liu et al. [15] disentangle and unify different forms of graph convolution into a multi-scale spatio-temporal operator, allowing flexible aggregation of local and long-range joint connectivity. Gao et al. [16] further show that learning heterogeneous spatial-temporal context helps to model long-range and cross-part interactions that are difficult to encode with a single, homogeneous graph. Building on these developments, graph-based methods tailored to hand tremor, such as the causality-aware framework by Quan et al. [17], treat joints as nodes in a dynamic graph and explicitly model causal relations that drive clinical scores across multi-centre data. However, current skeleton-based methods still depend on accurate pose estimation, so occlusions and low resolution can introduce keypoint jitter and distort tremor cues. Their sparse joint representation also misses fine finger micro-movements and may adapt poorly to real-world variations in lighting, cameras and patient behaviour.

B. RGB-Based Methods

RGB-based methods work directly on video data, without first going through pose estimation; in principle this means that these approaches maintain more information related to fine tremor detail. Early work showed that a normal smartphone camera plus Eulerian Video Magnification (EVM) can reveal tremor that is almost invisible to the naked eye [18], and studies on datasets like TIM-Tremor found that both hand-trajectory (Lagrangian) and phase/intensity (Eulerian) signals can reliably recover the main tremor frequency from RGB video [19] [20]. More recently, iris-based self-calibration and deep 3D hand tracking have been used to obtain both postural tremor amplitude and frequency from uncalibrated clinical videos, eliminating the need for physical markers or rulers and moving closer to real-world clinical workflows [21]. On top of these signal-processing pipelines, deep video models now operate directly on magnified RGB sequences [10] or pixel-level spatiotemporal volumes [22], using temporal-difference shift modules or 3D convolutional recurrent networks to regress MDS-UPDRS tremor subscores and overall severity from hand videos [10], [22]. These methods bypass pose estimation, capture finer motion details, enhance robustness to environmental variations. However, current RGB-based methods remain sensitive to lighting changes, compression artefacts, and camera shake. To address these issues, this paper proposes a novel tremor assessment framework based on the motion magnitude derived from optical flow components. By leveraging direction-invariant motion features instead of raw optical flow or pose estimation, the proposed method improves robustness to viewpoint changes and noise, while better aligning with common clinical evaluation criteria for hand tremor.

III. METHODS

A. Dataset and preprocessing

We recruited 65 participants with pathological hand tremor from Leeds Teaching Hospitals Trust. The study was performed in accordance with the ethical principles that have their origin in the Declaration of Helsinki. The study protocol was approved by the Yorkshire & The Humber - Sheffield Research Ethics Committee (REC) and HRA Approval was obtained (ref: 22/YH/0251). Informed consent was obtained from all participants prior to data collection. Each participant underwent the resting hand tremor assessment following Item 3.17 of the Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS). During recording, the camera was fixed at a stationary position directly in front of the participant's hands to ensure consistent spatial configuration across subjects. All videos were recorded with a spatial resolution of 1920×1280 pixels. The frame rate was set to 60 fps; on three occasions, the frame rate was set lower (30 fps, 15 fps) due to human error. Each recording lasted longer than 10 seconds and was stored in MP4 format. For each participant, both the left and right hands were recorded separately, resulting in two resting tremor videos per subject.

Two specialist neurologists who had completed standardized MDS-UPDRS training independently evaluated the recorded videos and graded the video on a score of 0 to 4, using the resting tremor criteria of the MDS-UPDRS. After independent grading, the neurologists consulted on any disagreements to provide a consensus grade that was used as the ground-truth label.

After data collection, preprocessing was performed to both standardize the input frame rate and augment the training data for subsequent model learning. Specifically, each 60 fps video was uniformly sampled to generate four new 15 fps videos, and each 30 fps video was uniformly sampled to generate two new 15 fps videos. Videos originally recorded at 15 fps were kept unchanged. Through this process, the dataset was expanded from the original 130 videos to 425 videos. Videos labelled with an MDS-UPDRS tremor score of 4 were grouped with those that scored three due to the small number of examples in the former category. During the subsequent training-testing split, all newly generated videos derived from the same high-frame-rate source video were strictly assigned to the same subset to avoid data leakage.

B. Motion-magnitude from optical flow

Compared with RGB representations, optical flow more effectively captures pixel-wise motion variations between consecutive frames, which aligns well with our objective of extracting tremor-related motion information from videos. For a PD tremor video with frame I_t , the optical flow (u, v) between consecutive frames is calculated as:

$$(u, v) = \arg \min_{u, v} \|I_{t+1}(\mathbf{x} + \mathbf{w}) - I_t(\mathbf{x})\|^2 + \lambda \|\nabla \mathbf{w}\|^2 \quad (1)$$

where $\mathbf{x} = (x, y)$ denotes the pixel coordinate and $\mathbf{w} = (u, v)$ represents the optical flow vector, with u and v corresponding to the horizontal and vertical flow components. The two-channel optical flow thus encodes directional motion information. However, in this application, the direction of tremor is not of primary interest. Therefore, to eliminate the additional channel and redundant directional information, we compute the motion magnitude M from u and v .

$$M = \sqrt{u^2 + v^2} \quad (2)$$

Downstream processing of data uses motion-magnitude rather than the original optical flow, such that the shape of input data changes from $(N \times 2 \times H \times W)$ to $(N \times H \times W)$. This simple change preserves tremor-related motion information in the video and halves the number of input channels, thereby decreasing the model input size and computational cost of downstream model training. In addition, the reduced input dimensionality lowers GPU memory consumption, allowing more video frames to be processed under the same hardware constraints.

C. Temporal difference

In our previous work, a temporal difference module was employed to effectively extract inter-frame variations. This

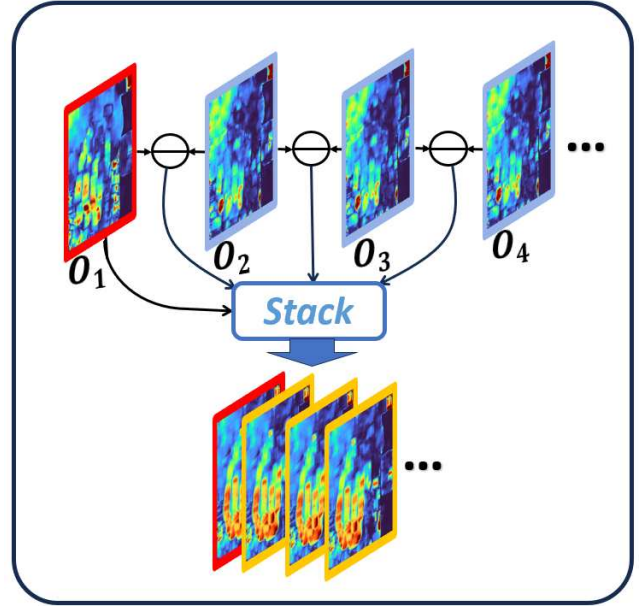


Fig. 1. Temporal differencing of motion magnitude: the magnitude of the first frame is preserved, and the differences between adjacent frames are computed and stacked along the temporal dimension.

simple structure enabled efficient modeling of motion changes between consecutive frames. Due to the structural constraints of the baseline model, temporal differencing was previously performed on groups of five consecutive frames. In the present work, we extend the temporal difference operation to the entire input frame sequence. The updated temporal difference formulation is illustrated in Fig. 1.

Specifically, given the original input frame sequence S :

$$S = [O_1, O_2, O_3, \dots, O_n] \quad (3)$$

where n denotes the total number of input frames and the input shape is $(N \times n \times H \times W)$, we perform frame-wise temporal differencing while retaining the first frame to preserve the initial information. The resulting sequence is defined as:

$$S' = [O_1, O_2 - O_1, O_3 - O_2, \dots, O_n - O_{n-1}] \quad (4)$$

Compared with our previous work, the current dataset was augmented to provide significantly more videos to enable more robust training. Notably, after the above processing, the overall input shape remains unchanged at $(N \times n \times H \times W)$.

D. Enhancing Convolutional Neural Network Backbones with Global Temporal Shift

We employ a Convolutional Neural Network (CNN) backbone to extract features from the motion magnitude obtained after temporal differencing. For the input frame sequence, instead of adopting the commonly used loose random sampling strategy in video classification, we utilize a dense sampling scheme with small temporal intervals. This strategy ensures temporal continuity of tremor motion while avoiding distortion. To accommodate conventional CNN backbones, the input

frame sequence is further divided into multiple segments, while preserving the original temporal continuity of the video. Given $S' = (S_1, S_2, \dots, S_N)$, the feature tensor of the i -th segment S_i after convolutional processing is defined as:

$$S_i = [C_{i1}, C_{i2}, \dots, C_{ic}], \quad (1 \leq i \leq N) \quad (5)$$

where c denotes the total number of feature channels. Since tremor assessment is inherently a global decision process, we propose a Global Shift Module (GSM) to enhance feature interaction across different temporal segments. Specifically, for each segment, a fixed proportion of channel features is selected and removed, and these features are uniformly redistributed across all segments. The newly obtained channel features are concatenated to replace the removed channels. The updated feature tensor of the i -th segment is thus formulated as:

$$S'_i = [C_{1i}, C_{2i}, \dots, C_{ji}, \dots, C_{ic}], \quad (1 \leq i \leq N) \quad (6)$$

where j denotes the number of exchanged feature channels. Through this simple operation, global feature exchange of tremor representations across temporal segments is achieved. The features of each segment at a given time step can interact with those from any other time step. Importantly, this interaction preserves the original feature tensor shape, without introducing additional channel dimensions or computational overhead. Another major advantage of this module is its transferability as a plug-and-play component, enabling seamless integration into different CNN backbones.

IV. EXPERIMENTS AND RESULTS

In this section, we present the performance of the proposed method on our dataset. We first compare the results of the motion-magnitude-based approach with those of the original optical flow-based approach. The comparison includes not only model performance but also computational cost and convergence efficiency. Based on this analysis, we further provide a comprehensive evaluation of the proposed method on our dataset.

A. Experiment settings and evaluation metrics

All experiments were implemented in PyTorch and conducted on a single NVIDIA GTX 5080 GPU. The model was optimized using cross-entropy loss and stochastic gradient descent. The best performance was obtained after 30 epochs, with an initial learning rate of 0.002, which was decayed by a factor of 0.1 at epochs 10 and 20. The batch size was set to 32, and the dropout rate was 0.5. ResNet-18 was adopted as the backbone network. In the proposed GSM, the proportion of globally shifted channels m was set to $1/4$. The number of video segments N was fixed at 18. The total training time for each experiment was approximately 0.5 hours.

To evaluate the performance of the proposed method, four evaluation metrics were reported for each experiment: average F1-score, area under the ROC curve (AUC), precision (Pre), recall (Rec), and *acceptable accuracy* (accuracy within ± 1 score tolerance). In addition, confusion matrices were provided to illustrate the model performance across different rating

TABLE I
COMPARISON OF OVERALL AND ACCEPTABLE ACCURACY BETWEEN OPTICAL FLOW AND MOTION MAGNITUDE REPRESENTATIONS.

Representation	Overall Acc. (%)	Acceptable Acc. (%)
Optical Flow	68.97	97.70
Motion Magnitude	71.76	98.35
Improvement	+2.79	+0.65

levels. Considering the limited dataset size, all evaluations in this study were conducted using participant-based 5-fold cross-validation. The training and test sets were in each fold were selected via stratified sampling to ensure that label distribution was consistent between folds. All reported metrics in this section were obtained from the test sets.

B. Comparison of Optical Flow Representations

In this experiment, we quantitatively compared the optical flow representation with the proposed motion magnitude representation. Table I summarized the comparison in terms of overall accuracy and acceptable accuracy ($|\Delta| \leq 1$). As shown in Table I, the motion magnitude representation achieved an overall accuracy of 71.76%, outperforming the optical flow baseline (68.97%) by 2.79%. Furthermore, the acceptable accuracy increased from 97.70% to 98.35%, corresponding to a 0.65% absolute improvement. These results indicated that replacing the two-channel directional optical flow with a direction-invariant motion magnitude representation yielded consistent performance gains in PD tremor severity assessment.

In addition to predictive performance, we further evaluated the computational efficiency of both representations. Fig. 2 presented the comparison in terms of computational complexity, memory consumption, and runtime. As illustrated in Fig. 2, adopting motion magnitude reduced the computational cost by 3.54 GFLOPs, decreased the reserved GPU memory by 18 MB, and shortened the runtime by 8 minutes, corresponding to a 21.1% reduction in execution time. These findings demonstrated that the proposed motion magnitude representation not only improved classification performance but also enhanced computational efficiency, thereby achieving a favorable trade-off between accuracy and efficiency.

C. Performance on Hand Tremor Assessment

In this section, we comprehensively evaluated the performance of the proposed method for Hand tremor assessment. Table II summarized the per-score evaluation metrics obtained through five-fold cross-validation, including precision, recall, F1-score, and score-wise accuracy. The results indicated that the proposed framework achieved competitive overall performance across tremor severity levels. Grade 3 attained the highest F1-score (90.48%) and recall (93.44%), while Grade 0 also demonstrated balanced performance across all metrics. In comparison, Grades 1 and 2 yielded relatively lower recall and F1-scores, reflecting increased classification difficulty for intermediate tremor severity levels.

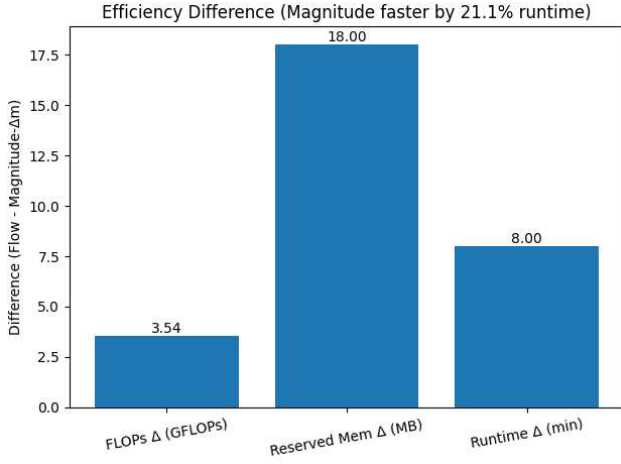


Fig. 2. Improvements in computational complexity, memory consumption and runtime in comparison to baseline optical flow method

TABLE II
PER-CLASS PERFORMANCE ON THE AGGREGATED 5-FOLD TEST SET.
OVERALL ACCURACY AND ACCEPTABLE ACCURACY ($|\Delta| \leq 1$) ARE REPORTED AT THE BOTTOM.

Score	Prec.	Rec.	F1	Class Acc.
0	70.06	86.11	77.26	82.82
1	74.53	50.97	60.54	75.76
2	58.44	69.23	63.38	87.76
3	87.69	93.44	90.48	97.18
Overall Accuracy			71.76%	
Acceptable Accuracy ($\Delta \leq 1$)			98.35%	

To further analyze the prediction behavior of the model, Fig. 3 presented the row-normalized confusion matrix aggregated over five folds. Unlike the metric-based summary in Table II, the confusion matrix revealed detailed misclassification patterns. It was observed that Grade 1 was frequently confused with Grade 0 (0.32), suggesting partial overlap between mild and moderate tremor presentations. Similarly, Grade 2 was occasionally misclassified as Grade 1 (0.14) or Grade 3 (0.11). In contrast, Grade 3 showed minimal confusion with other scores, indicating clearer discriminative boundaries at the severe end of the spectrum. The ROC analysis further evaluated the discriminative capability of the proposed method. The AUC values indicated strong separability for Grades 0 and 3, whereas intermediate scores exhibited relatively lower discrimination performance. This observation was consistent with the confusion patterns shown in Fig. 4. As shown in Table III, most prediction errors occur within adjacent tremor severity levels. Specifically, 98.35% of predictions differ from the ground truth by no more than one score level, resulting in a quadratic weighted kappa (QWK) of 0.853, indicating strong agreement between the predicted and clinical scores. Overall, these results demonstrated that the proposed model provided stable and effective performance for automatic PD tremor severity assessment.

TABLE III
ERROR DISTANCE DISTRIBUTION OF TREMOR SCORE PREDICTIONS ON THE AGGREGATED TEST SET.

Error Distance $ y - \hat{y} $	Ratio (%)
0	71.76
1	26.59
2	2.35
3	0.00
Quadratic Weighted Kappa (QWK)	0.853

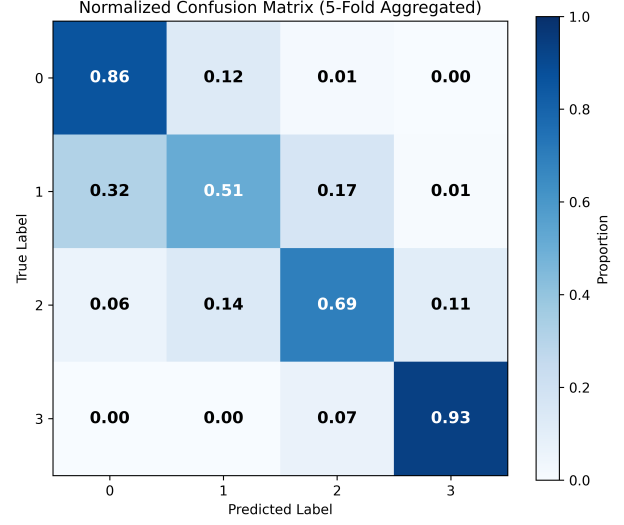


Fig. 3. Aggregated confusion matrix over 5 cross-validation, illustrating the classification performance across all tremor severity levels.

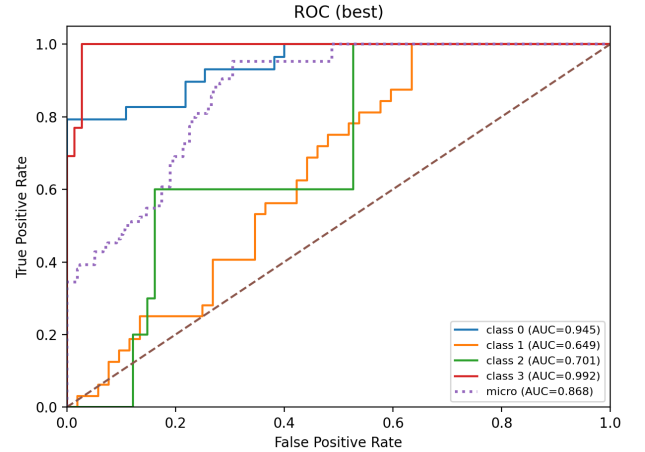


Fig. 4. ROC curves of the proposed method for four tremor severity classes. The micro-average ROC curve is also reported.

V. CONCLUSION

In this work, we proposed a motion magnitude framework for PD tremor videos assessment as an alternative to optical flow-based approaches. The motion magnitude is derived from the two directional components of optical flow, which suppresses redundant directional noise while reduc-

ing the input data dimensionality. We further extended the temporal differencing operation to the entire frame sequence and adopted a CNN backbone embedded with the proposed Global Shift Module to facilitate feature interaction across all temporal segments. This design enables effective extraction of tremor-related features for predicting MDS-UPDRS tremor scores from videos. Experimental results demonstrated that the motion-magnitude representation outperforms the original two-channel optical flow in terms of accuracy, memory efficiency, and convergence stability. Moreover, evaluations on a relatively large-scale dataset indicate that the proposed method provides an efficient and effective solution for automatic quantification of hand tremor severity from video data. Nevertheless, several limitations should be noted. Although the dataset used in this study is clinically meaningful, its size remains relatively limited. In addition, merging Grades 3 and 4 helps address class imbalance but reduces the granularity of severe tremor assessment. Finally, the current framework still relies on relatively complex models. In future work, we plan to explore more efficient and lightweight architectures, improve robustness under diverse recording conditions, and validate the proposed method on broader datasets to further enhance its clinical applicability.

REFERENCES

- [1] R. J. Elble, "The essential tremor rating assessment scale," *Journal of Neurology & Neuromedicine*, vol. 1, no. 4, 2016.
- [2] C. G. Goetz, B. C. Tilley, S. R. Shaftman, G. T. Stebbins, S. Fahn, P. Martinez-Martin, W. Poewe, C. Sampaio, M. B. Stern, R. Dodel *et al.*, "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*, vol. 23, no. 15, pp. 2129–2170, 2008.
- [3] N. P. Bajaj, V. Gontu, J. Birchall, J. Patterson, D. G. Grosset, and A. J. Lees, "Accuracy of clinical diagnosis in tremulous parkinsonian patients: a blinded video study," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 81, no. 11, pp. 1223–1228, 2010.
- [4] D. C. Wong and S. Williams, "Artificial intelligence analysis of videos to augment clinical assessment: an overview," *Neural Regeneration Research*, vol. 19, no. 4, pp. 717–718, 2024.
- [5] M. U. Friedrich, S. Relton, D. Wong, and J. Alty, "Computer vision in clinical neurology: A review," *JAMA neurology*, vol. 82, no. 4, pp. 407–415, 2025.
- [6] F. Zhang, V. Bazarevsky, A. Vakunov, A. Tkachenka, G. Sung, C.-L. Chang, and M. Grundmann, "Mediapipe hands: On-device real-time hand tracking," 2020.
- [7] V. Filippou, A. Alty, S. Relton, S. Williams, K. Bliss, and D. Wong, "Comparison of monocular video-based methods for measuring the amplitude of hand tremor," *Computers in Biology and Medicine*, in press, 2026.
- [8] L.-R. Müller, J. Petersen, A. Yamlahi, P. Wise, T. J. Adler, A. Seitel, K.-F. Kowalewski, B. Müller, H. Kenngott, F. Nickel *et al.*, "Robust hand tracking for surgical telestration," *International journal of computer assisted radiology and surgery*, vol. 17, no. 8, pp. 1477–1486, 2022.
- [9] S. Williams, H. Fang, S. D. Relton, D. C. Wong, T. Alam, and J. E. Alty, "Accuracy of smartphone video for contactless measurement of hand tremor frequency," *Movement Disorders Clinical Practice*, vol. 8, no. 1, pp. 69–75, 2021.
- [10] W. Liu, X. Lin, X. Chen, Q. Wang, X. Wang, B. Yang, N. Cai, R. Chen, G. Chen, and Y. Lin, "Vision-based estimation of mds-updrs scores for quantifying parkinson's disease tremor severity," *Medical Image Analysis*, vol. 86, p. 102754, 2023.
- [11] J. Amo-Salas, A. Olivares-Gil, Á. García-Bustillo, D. García-García, Á. Arnaiz-González, and E. Cubo, "Computer vision for parkinson's disease evaluation: A survey on finger tapping," *Healthcare*, vol. 12, no. 4, p. 439, 2024.
- [12] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "Openpose: Realtime multi-person 2d pose estimation using part affinity fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 172–186, 2019.
- [13] H. Zhang, E. S. Ho, F. X. Zhang, and H. P. Shum, "Pose-based tremor classification for parkinson's disease diagnosis from video," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2022, pp. 489–499.
- [14] H. Zhang, E. S. Ho, F. X. Zhang, S. Del Din, and H. P. Shum, "Pose-based tremor type and level analysis for parkinson's disease from video," *International Journal of Computer Assisted Radiology and Surgery*, vol. 19, no. 5, pp. 831–840, 2024.
- [15] Z. Liu, H. Zhang, Z. Chen, Z. Wang, and W. Ouyang, "Disentangling and unifying graph convolutions for skeleton-based action recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2020, pp. 143–152.
- [16] X. Gao, Y. Yang, Y. Wu, and S. Du, "Learning heterogeneous spatial-temporal context for skeleton-based action recognition," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 9, pp. 12 130–12 141, 2024.
- [17] Y. Quan, R. Guo, J. Ren, C. Zhang, and X. Qian, "Discovering causality representations in graph convolutional network for automated assessment of parkinsonian hand postural tremor," *IEEE Transactions on Neural Networks and Learning Systems*, vol. PP, no. 99, pp. 1–14, 2026.
- [18] S. Williams, H. Fang, J. Alty, R. Qahwaji, P. Patel, and C. D. Graham, "A smartphone camera reveals an 'invisible' parkinsonian tremor: a potential pre-motor biomarker?" *Journal of Neurology*, vol. 265, no. 12, pp. 3017–3018, 2018.
- [19] S. L. Pinteá, J. Zheng, X. Li, P. J. Bank, J. J. van Hilten, and J. C. van Gemert, "Hand-tremor frequency estimation in videos," in *Proceedings of the European conference on computer vision (ECCV) workshops*, 2018, pp. 0–0.
- [20] G. Güney, T. S. Jansen, S. Dill, J. B. Schulz, M. Dafotakis, C. H. Antink, and A. K. Braczynski, "Video-based hand movement analysis of parkinson patients before and after medication using high-frame-rate videos and mediapipe," *Sensors*, vol. 22, no. 20, p. 7992, 2022.
- [21] D. L. Guarín, "Video-based quantification of hand postural tremor without external references. integrating postural tremor quantification into visionmd," *npj Parkinson's Disease*, vol. 11, no. 1, p. 351, 2025.
- [22] F. Duque-Quiceno, G. Sarapata, Y. Dushin, M. Allen, and J. O'Keeffe, "Deep learning for objective estimation of parkinsonian tremor severity," 2024.