Improving Automatic Tremor and Movement Motor Disorder Severity Assessment for Parkinson's Disease with Deep Joint Training

Chun-Min Chang[1,2], Yu-Lin Huang[1,2], Jui-Cheng Chen[3], Chi-Chun Lee[1,2]

Abstract— **Parkinson's disease (PD) is one of the most severe and common disease globally. PD induces motor system impairment causing symptoms such as shaking, rigidity, slowness of movement, body tremor and difficulty with walking. Clinically, accurately and objectively assessing the severity of PD symptoms is critical in controlling appropriate dosage of Levodopa to prevent unwanted side effect of switching between Dyskinesia and PD. The unified Parkinson's disease rating scale published by the Movement Disorder Society (MDS-UPDRS) is an validated instrument regularly administrated by trained physician to assess the severity of a PD patient's motor disorder. In this work, we aim at advancing vision-based automatic motor disorder assessment, specifically hand tremor and movement, for PD patients during UPDRS. Our proposed method leverages information across the two behavior tasks simultaneously via deep joint training to improve each single task's, i.e., tremor and movement, severity classification rate. We evaluate our framework on a large cohort of 106 PD patients, and with our proposed deep joint training framework, we achieve accuracy of 78.01% and 80.60% in right and left hand movement binary classification; in terms of tremor severity classification, our approach obtains an enhanced recognition rates of 72.20% and 71.10% for right and left hand respectively.**

I. INTRODUCTION

Parkinson's disease (PD) is a severe long-term degenerative nervous system disorder. In 2015, PD happened to 6.2 million people causing 117 thousand death. To date, there remains no cure for Parkinson's disease. Most PD patients suffer from motor system disorder, e.g., shaking, rigidity, slowness of movement, body tremor, and difficulty with walking [1]. Levodopa is regularly prescribed to PD patients in order to ease these symptoms [2]. However, the long-term overdosing of Levodopa may cause a side effect called Dyskinesia. Specifically, this side effect creates an "On-and-Off" state. When too much Levodopa is taken, "On" state would be activated, and Dyskinesia would show; too little Levodopa would results in an "Off" state where Parkinsonism would show. In current clinical practices, identifying a proper balance in the Levadopa dosage for each PD patient is one of the most critical issue in the treatment course.

Clinically, physicians use the unified Parkinson's disease rating scale published by the Movement Disorder Society (MDS-UPDRS) to evaluate the severity of the disordered symptoms in order to assist in prescribing medications for PD patients [3]. UPDRS is usually performed in a clinical setting relying on expert's manual ratings. This restriction prevents the UPDRS to be administrated for scalable applications, e.g., in home and/or assessment on demand. This assessment can also be susceptible to inter-physician variability. Developing objective methods in assessing the severity of impaired motor symptoms is critical for PD patient's healthcare.

Most of the previous automatic PD assessment research has actively investigated the use of wearable sensors. For example, Taewoong et al. performed automatic PD motor-state recognition with a wrist-worn wearable sensor using deep learning [4]; Lang et al. used multi-layer Gaussian Process to estimate motor symptoms severity [5]. Giuberti et al. instrumented patients with wireless body-worn inertial nodes in order to estimate the relationship between UPDRS score and leg agility task [6], and Hssayeni et al. developed a mediation state detection for Parkinson's disease also using wearable sensors [7]. However, wearable solutions are generally expensive and require extensive effort in instrumenting patients. Recently, researchers have moved toward vision-based PD symptoms assessment. For example, Rao et al. validated a new score for video-based Dyskinesia severity [8]. Li et al. developed a vision-based method to detect motor system disorder using pose estimation [9] and further monitored patients after Levodopa infusions using video camera for a 9 PD patient cohort [10]. Pintea estimated the hand tremor of PD also from the videos [11].

While several works have started to investigate vision-based approaches for automatic PD symptom assessment, they only concentrate on a single task (e.g., hand tremor only) and on a very limited number of patient samples. In this work, we collect a database of 106 PD patients recordings during UPDRS interviews. PD is a nervous system degenerative disorder affecting motor system with symptoms potentially manifested simultaneously across different UPDRS tasks. Hence, we propose to perform automatic hand tremor and movement disorder severity assessment (the two major contributing factors of UPDRS) through deep joint training to improve the accuracy of each single task. Specifically, our method obtains accuracy of 78.01% and 80.60% in right and left hand movement's binary classification task; in terms of tremor severity classification, our proposed approach obtains 72.20% (right) and 71.10% (left) accuracy.

II. METHODOLOGY

A. Database

Our database is collected at the Department of Neurology, China Medical University (CMUH), Taiwan.[1]It consists of

[1]CMC, YLH, CCL are with Department of Electrical Engineering, National Tsing Hua University, Taiwan cclee@ee.nthu.edu.tw

[2]CMC, YLH, CCL are with MOST Joint Research Center for AI Technology and All Vista Healthcare, Taiwan

[3]JCC is with Department of Neurology, China Medical University, Taiwan

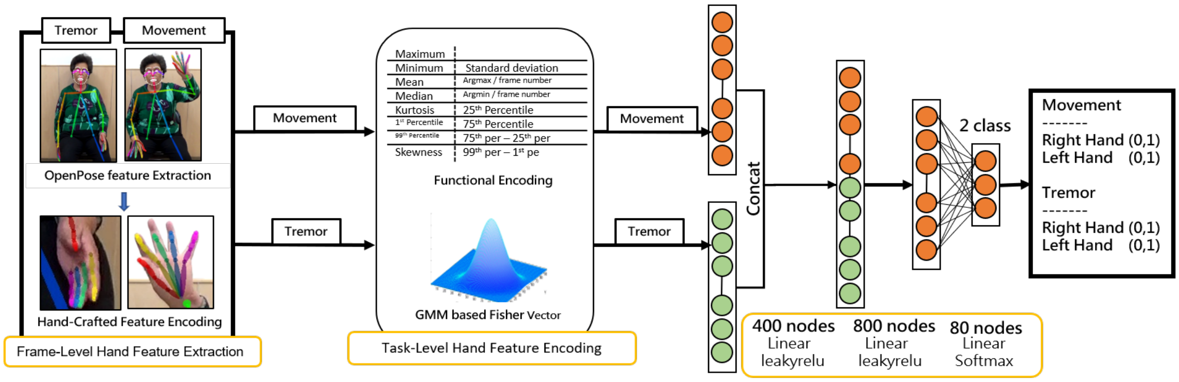

Fig. 1. Figure shows our proposed framework. OpenPose tool kit is used to to extract key points from hand, each hand is then encoded using task-level encoding method. Task 2 and Task 7 labels are predicted from concatenating the two extracted hand features simultaneously via joint training

106 PD patients performing the 13 tasks derived from instrumentation of UPDRS (approximately 12 minutes total). Each task is marked on a scale between [0, 4] by neurologist watching video as a measurement on the severity of each PD symptom. In this work, we focus first on the two major hand controlling tasks (Task 2 and Task 7). Task 2 is the Movement task (a.k.a.,"Finger Taps") that requires the patient to tap thumb and index finger together in rapid succession. Task 7 is the Tremor task, where doctor checks the patient's tremor state when at relax. Video recording is done with Resolution: 1280x720, 30 FPS using a three camera setup placed in three different positions: center (SONY HDR-PJ675), right and left (SONY HDR-CX405) (sample collection is shown in Figure 2). We further binarize the movement assessment score for each hand into two classes (a score of 0 means normal and scores between [1 - 4] as abnormal) and the tremor assessment score is divided into two classes ([0 - 1] means normal and [2 - 4] as abnormal) to conduct our classification task.

B. Frame-Level Hand Feature Extraction

We utilize the OpenPose toolkit, i.e., a real-time multi-person key points (including face, hands and body postures) detection library [12], to track the key points of the hand at 30 frames per second. An illustration on the key points tracked is shown in Figure 3. To further construct our frame-level hand features from the recordings, we first take the X-coordinates, its difference in time (Δx) and the Y-coordinates with its difference in time (Δy) as the moving distance of the nine points (point 0 to 8). Then, we group eight points into pairs (point 1 with point 5, point 2 with point 6, point 3 with point 7, and point 4 with point 8) and calculate the distance, the velocity, and the acceleration between the points within each pair. We concatenate the 48 values as our frame-level hand movement feature vector. Specifically, we compute this frame-level feature vector on the center position camera because it generally captures the patient's hand much clearer without occlusion. We extract 10 seconds of each task separately resulting in a 300 frames of frame-level hand features used in this work.

[1]Our study is approved by the Institution of Review Board of Medical University, IRB#:CMUH105-REC2-055

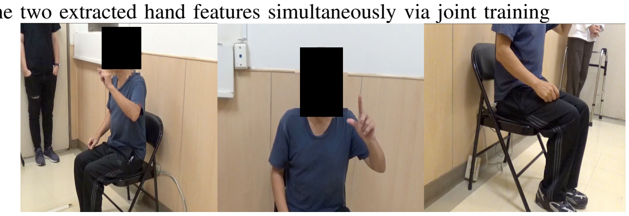

Fig. 2. It shows our camera setup in three positions: center, right, and left.

C. Task-Level Hand Feature Encoding

Our aim is to automatically classify two classes (normal vs. abnormal) for each hand at the individual task level (Task 2 and 7). We additionally encode these 300 frames of frame-level features to a single task-level feature vector using two different encoding approaches: functional encoding, and Fisher vector encoding.

1) Functional: In this work, we compute 15 statistical functions, i.e., max, min, mean, median, standard deviation, 1st percentile, 99th percentile, 99th percentile−1st percentile, skewness, kurtosis, minimum position, maximum position, lower quartile, upper quartile,interquartile range, over the 300 frame-level features resulting in an encoded vector with dimension 720.

2) Fisher vector (FV): We also perform Fisher-vector encoding as another approach in deriving task-level feature. FV has been used extensively in computer vision applications [13]. FV encoding first assumes the data is generated from a Gaussian Mixture Model (GMM), and in order to represent a data sample X, we define a scoring function as below:

$$G_\lambda^X = \nabla_\lambda log u_\lambda(X)$$

where $u_\lambda(X)$ denotes the likelihood of X given the probability distribution function (PDF) of a GMM. λ represents the parameters of GMM, $\lambda = w_k, u_k, \sum_k, k = 1, ..., K$. G_λ^X is the direction where λ has to move to provide a better fit between u_λ and X. Fisher vector is derived by computing the following first and second order statistics:

$$g_{u_k}^X = \frac{1}{T\sqrt{w_k}} \sum_{t=1}^{T} \gamma_t(k) \left(\frac{x_t - u_k}{\sigma_k} \right)$$

$$g_{\sigma_k}^X = \frac{1}{T\sqrt{2w_k}} \sum_{t=1}^{T} \gamma_t(k) \left(\frac{(x_t - u_k)^2}{\sigma_k^2} - 1 \right)$$

TABLE I

SUMMARY OF CLASSIFICATION RESULTS FOR LEFT AND RIGHT HAND IN THE TREMOR AND THE MOVEMENT SEVERITY TASK

	Results									
	Functi	FV-4	FV-8	FV-16	FV-32	Functi	FV-4	FV-8	FV-16	FV-32
Movement	Right Hand					Left Hand				
Concat-DNN	66.41	76.80	**78.01**	72.40	69.70	73.50	**80.60**	80.28	78.34	79.40
Single-DNN	64.51	71.95	71.08	72.04	66.77	72.84	73.53	76.11	75.93	72.99
Concat-SVM	72.30	71.20	72.10	67.60	67.40	67.80	75.00	78.80	76.10	72.30
Single-SVM	69.70	69.50	65.90	66.00	66.00	75.20	71.00	76.40	70.30	73.10
Tremor	Right Hand					Left Hand				
Concat-DNN	**72.20**	71.10	65.24	64.08	64.71	**71.10**	65.70	65.04	67.14	67.02
Single-DNN	62.83	64.08	56.70	54.62	53.07	60.04	50.28	50.00	54.60	52.30
Concat-SVM	69.30	66.90	63.30	67.10	59.40	57.80	62.10	55.70	57.90	62.20
Single-SVM	62.20	64.70	63.20	64.50	63.10	59.80	59.10	59.80	63.40	63.50

$\gamma_t(k)$ is defined as

$$\gamma_t(k) = \frac{w_k u_k(x_t)}{\sum_{j=1}^{K} w_j u_j(x_t)}$$

where $w_k, u_k, \sum_k, k = 1, ..., K$ correspond to mixture weight, mean, and covariance matrix for each mixture of Gaussian. In specifics, we use GMMs with different number of mixture (4-GMM, 8-GMM, 16-GMM, 32-GMM) and retrieve the mean and variance as our final FV-encoded task-level features for each data sample (dimension = $48 \times 2 \times mixture\_number$).

D. Deep Joint Training

Figure 1 shows the complete structure of our proposed framework. The encoded features from the two hand-controlling tasks (Task 2 and Task 7) are concatenated to be fed into a single deep feedforward neural network that classify individual label of movement (left and right) and tremor (left and right). Let's denote the two input features: movement feature, X_m and tremor feature, X_t where $X_m = [x_m^1, x_m^2, ..., x_m^N]$ and $X_t = [x_t^1, x_t^2, ..., x_t^N]$ and total sample number N. We utilize the following joint loss function, L, to optimize our network:

$$K = \sigma W_3(\sigma W_2(W_{m1}(X_m), W_{t1}(X_t)))$$

$$L = -Y \log K + (1 - Y) \log(1 - K)$$

where Y indicates the label of interest (two hands) for the two tasks, Y_{mR}, Y_{mL}, Y_{tR}, Y_{tL}, where they all contain $[y^1, y^2, ..., y^N]$, and W_{m1}, W_{t1}, W_2, W_3 represents the parameters of each layer. σ is the activation function. The use of joint training aims at leveraging the hand motor behaviors that each patient exhibits during Task 2 and 7 together to improve each of the two tasks classification rate.

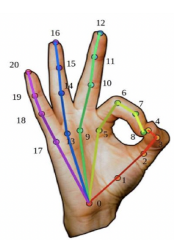

Fig. 3. An illustration shows the key points of hand extracted from the OpenPose. [12]

III. EXPERIMENTAL SETUP AND RESULT

We compare our methods with several baselines. The first is the standard baselines, i.e., each extracted task-level features are trained using Support-Vector Machine (SVM) or DNN separately without joint training (Single-SVM, Single-DNN). Second, we use SVM for each label by feeding with concatenated (Task 2 and Task 7) features together (Concat-SVM). Finally our method is based on joint training on concatenated task-level features (Concat-SVM). We perform leave-one-patient out cross validation with the performance metric of unweighted average recall (UAR).

A. Result and Discussion

Table I summarizes our results. For Single-SVM, the best recognition rates obtained are 69.7% and 76.4% respectively for right and left hand in the Movement task, and the best recognition rates are 64.7% and 63.50% for right and left hand classification in the Tremor task. Simply concatenating features of both tasks (Concat-SVM), which introduces behavior information from the other task to the current task, the best classification accuracy are 72.30% (increases about 2.6%) and 78.80% (increases about 2.6%) for right and left hand classification in the Movement task. For the Tremor task, the recognition rate for the right hand also improves to 69.30%, i.e., an increase of 4.6%.

We observe that through the use of deep joint training (Concat-DNN) that we propose, the classification result of each task reaches 78.01%, 80.60%, 72.20% and 71.10%, which improves over the baseline Single-DNN about 6%, 4%, 8% and 11% respectively on right and left hand recognition in the Movement task. The right and left hand recognition in the Tremor task improves over the Single-SVM method at around 9%, 4%, 8% and 7% respectively. The proposed Concat-DNN is the best performing model, which achieves accuracy of 78.01% and 80.60% in right and left hand classification accuracy in the Movement task; in terms of tremor severity classification, our approach obtains recognition rates of 72.20% and 71.10% for right and left hand respectively.

B. Movement and Tremor Analysis

We carry out a further analysis to understand which particular cross-task behaviors may be contributing in the improvement of classification rates for our proposed deep joint training framework. We first compute cross-task behavior-to-label Spearman correlations, i.e., a patient's Tremor's task-level hand features to his/her Movement assessment score,

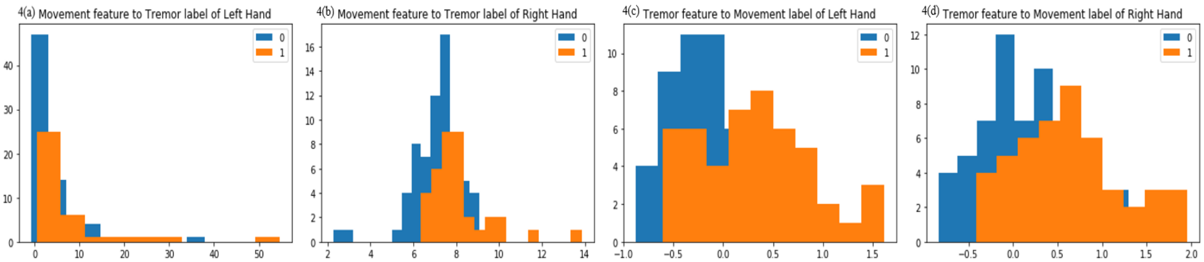

Fig. 4. Histograms on distribution of Tremor behaviors to Movement label or vice versa for normal (blue colored) vs. abnormal (orange colored).

and vice versa. We observe that moving distance of thumb during the Tremor task is correlated highly to the Movement score of a patient. On the other hand, the distance between thumb and index finger (and their associated velocity and acceleration) during the Movement task is correlated to the Tremor score of a patient.

We further present a histogram demonstrating the discriminability of the cross-task behaviors (Task 2 behaviors for Task 7 label and vice versa) between the normal vs. abnormal PD patients. Orange color indicates the patients with disorder of tremor or movement (abnormal). Blue is the patients without such disorder (normal). The plot of 4(a) shows distribution of the moving distance of thumb between the two groups; the plot of 4(b) shows the distribution of the distance of [the head of thumb, the head of index finger] between the two groups; the plots of 4(c) and 4(d) show the distribution of distance of [the root of thumb, the root of index finger] between the two groups.

In these four plots, we observe a distinct distributions between abnormal (Orange) vs. normal (Blue). Generally, we see that moving distance of thumb from in the Movement task are related to the Tremor score, and the distance between thumb and index finger during the Tremor task is related to the Movement score. The complementary information may have resulted from the fact that PD patient has a central motor system disorder that affects his/her hand muscle movements jointly even when being assessed to perform different targeted actions. This phenomenon, as captured by our deep joint training framework, are being leveraged in advancing the automatic assessment accuracy.

IV. Conclusions

In this work, we propose a deep jointly training network to perform vision-based automatic classification between normal vs. abnormal hand-related Tremor and Movement states for PD patients during UPDRS assessment. Specifically, the joint training framework leverages the cross-tasks behaviors exhibited by the patient to improve the classification rate of the target task. An immediate future work is to integrate all 13 tasks in the UPDRS to move toward a fully clinically-relevant automatic severity assessment for PD patients. At the same time, we would like to investigate how does a PD patient's core motor system affects each type of targeted muscle-controlling actions (e.g. rapid movement, stable controlling, precise touching, etc) in a large scale manner by continuously developing advanced vision-based deep learning algorithms to be used and validated clinically.

References

[1] C Robertson and KA Flowers, "Motor set in parkinson's disease.," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 53, no. 7, pp. 583–592, 1990.

[2] Claire L Tomlinson, Rebecca Stowe, Smitaa Patel, Caroline Rick, Richard Gray, and Carl E Clarke, "Systematic review of levodopa dose equivalency reporting in parkinson's disease," *Movement disorders*, vol. 25, no. 15, pp. 2649–2653, 2010.

[3] Christopher G Goetz, Barbara C Tilley, Stephanie R Shaftman, Glenn T Stebbins, Stanley Fahn, Pablo Martinez-Martin, Werner Poewe, Cristina Sampaio, Matthew B Stern, Richard 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.

[4] Terry Taewoong Um, Franz Michael Josef Pfister, Daniel Christian Pichler, Satoshi Endo, Muriel Lang, Sandra Hirche, Urban Fietzek, and Dana Kulic, "Parkinson's disease assessment from a wrist-worn wearable sensor in free-living conditions: Deep ensemble learning and visualization," *arXiv preprint arXiv:1808.02870*, 2018.

[5] Muriel Lang, Urban Fietzek, Jakob Fröhner, Franz MJ Pfister, Daniel Pichler, Kian Abedinpour, Terry T Um, Dana Kulić, Satoshi Endo, and Sandra Hirche, "A multi-layer gaussian process for motor symptom estimation in people with parkinson's disease," *arXiv preprint arXiv:1808.10663*, 2018.

[6] Matteo Giuberti, Gianluigi Ferrari, Laura Contin, Veronica Cimolin, Corrado Azzaro, Giovanni Albani, and Alessandro Mauro, "Assigning updrs scores in the leg agility task of parkinsonians: Can it be done through bsn-based kinematic variables?," *IEEE Internet of Things Journal*, vol. 2, no. 1, pp. 41–51, 2015.

[7] Murtadha D Hssayeni, Michelle A Burack, Joohi Jimenez-Shahed, Behnaz Ghoraani, et al., "Wearable-based mediation state detection in individuals with parkinson's disease," *arXiv preprint arXiv:1809.06973*, 2018.

[8] Anusha Sathyanarayanan Rao, Benoit M Dawant, Robert E Bodenheimer, Rui Li, John Fang, Fenna Phibbs, Peter Hedera, and Thomas Davis, "Validating an objective video-based dyskinesia severity score in parkinson's disease patients," *Parkinsonism & related disorders*, vol. 19, no. 2, pp. 232–237, 2013.

[9] Michael H Li, Tiago A Mestre, Susan H Fox, and Babak Taati, "Vision-based assessment of parkinsonism and levodopa-induced dyskinesia with pose estimation," *Journal of neuroengineering and rehabilitation*, vol. 15, no. 1, pp. 97, 2018.

[10] Michael H Li, Tiago A Mestre, Susan H Fox, and Babak Taati, "Automated assessment of levodopa-induced dyskinesia: Evaluating the responsiveness of video-based features," *Parkinsonism & related disorders*, vol. 53, pp. 42–45, 2018.

[11] Silvia L Pintea, Jian Zheng, Xilin Li, Paulina JM Bank, Jacobus J van Hilten, and Jan C van Gemert, "Hand-tremor frequency estimation in videos," in *European Conference on Computer Vision*. Springer, 2018, pp. 213–228.

[12] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 7291–7299.

[13] Ken Chatfield, Victor S Lempitsky, Andrea Vedaldi, and Andrew Zisserman, "The devil is in the details: an evaluation of recent feature encoding methods.," in *BMVC*, 2011, vol. 2, p. 8.