A Novel Method for Assessing the Severity of Levodopa-Induced Dyskinesia using Wearable Sensors

Sunghoon Ivan Lee, *Member*, *IEEE*, Jean-Francois Daneault, Fatemeh Noushin Golabchi, Shyamal Patel, *Member IEEE*, Sabrina Paganoni, Ludy Shih, Paolo Bonato, *Senior Member*, *IEEE*

***Abstract* - Patients with Parkinson's disease often experience significant changes in the severity of dyskinesia when they undergo titration of their medications. Dyskinesia is marked by involuntary jerking movements that occur randomly in a burst-like fashion. The burst-like nature of such movements makes it difficult to estimate the clinical scores of severity of dyskinesia using wearable sensors. Clinical observations are generally made over intervals of 15-30 s. On the other hand, techniques designed to estimate the severity of dyskinesia based on the analysis of wearable sensor data typically use data segments of approximately 5 s. Consequently, some data segments might include dyskinetic movements, whereas others might not. Herein, we propose a novel method suitable to automatically select data segments from the training dataset that are marked by dyskinetic movements. The proposed method also aggregates results derived from the testing dataset using a machine learning algorithm to estimate the severity of dyskinesia from wearable sensor data. Results obtained from the analysis of sensor data collected from seven subjects with Parkinson's disease showed a marked improvement in the accuracy of the estimation of clinical scores of dyskinesia.**

I. Introduction

Parkinson's disease (PD) is a progressive neurodegenerative disorder that is characterized by a loss of dopaminergic neurons that leads to tremor, muscle rigidity, slowness of movement (bradykinesia), and postural instability. As the disease progresses, these motor symptoms severely hamper the performance of most activities of daily living. The clinical management of motor symptoms in PD is typically based on the prescription of levodopa, which is a precursor of dopamine. However, long-term use of levodopa can lead to dyskinesia. The most common type of levodopa-induced dyskinesia (LID) is choreic peak-dose dyskinesia, which is marked by involuntary movements that are random, non-rhythmic in appearance as well as unsustained [1].

This work was partially supported by the Michael J. Fox Foundation under the project entitled "Wearable Sensors and a Web-Based Application to Monitor Patients with Parkinson's Disease in the Home Environment" and by the National Institutes of Health under the project enttiled "Wearable Wireless Sensors to Monitor Parkinsonian Symptoms" (#5R44AG029036-03).

S. I. Lee, J. F. Daneault, F. N. Golabchi, S. Patel, S. Paganoni and P. Bonato are with the Department of Physical Medicine and Rehabilitation, Harvard Medical School, Spaulding Rehabilitation Hospital, Boston, MA 02114 USA (e-mail: {slee117, jdaneault, fgolabchi, spatel19, spaganoni, pbonato}@partners.org)

L. Shih is with the Department of Neurology, BIDMC, Boston, MA 02215 USA (e-mail: lshih@caregroup.harvard.edu).

Conventional methods for assessing LID range from clinical scales to laboratory methods [2]. Clinical scales are used to assess the severity of LID as well as their impact on the performance of activities of daily living. They rely on the visual observation of the subject's motor behavior and rate the severity of LID based on their worst manifestation during an interval of typically of 15-30 s. Their reliance on ordinal scales limits their resolution. Besides, LID-specific clinical scales are affected by modest inter-rater reliability. Motion-tracking systems and force platforms have also been used to evaluate LID but these systems are expensive and their use is limited to the laboratory setting.

Recent advances in wearable technology appear to provide the tools needed to overcome the limitations of previously used techniques to assess the severity of LID. Wearable sensors allow one to acquire data in an unobtrusive manner from several hours to even days at relatively low cost. Several research groups have tackled the task of evaluating LID using wearable sensors. To date, wearable sensors that monitor acceleration [3], angular velocity [4], and electromyographic activity [5, 6] have been utilized to collect data and estimate the severity of LID. Cole et al [5] introduced an automated algorithm to detect LID and tremor using accelerometer, surface EMG data, and classifiers such as support vector machines and dynamic neural networks. In a subsequent paper [6], these authors used the same sensor data to estimate the severity of LID as mild, moderate, and severe. Tsipouras et al [4] used data recorded from accelerometer and gyroscope sensors to distinguish LID from other Parkinsonian symptoms, and estimated a whole-body dyskinesia severity index using different classification methods.

Our own research team [3] introduced a methodology based on accelerometer data to estimate the severity of LID as rated by the clinician during the performance of standardized tasks. We used a supervised method to assess the severity of LID in overlapping data segments of 5 s. Each data segment was considered as an individual data point. However, clinical scores were derived via observation of 30 s and captured the most severe manifestation of LID during that interval. Because dyskinetic movements occur randomly in a burst-like fashion, some data segments included them while others did not. Hence, both the training and testing datasets contained data segments that were not properly labeled, i.e. they were associated with high-severity score but included no dyskinetic movements or dyskinetic movements of milder severity.

In this paper, we propose a method to address the potential mismatch between the characteristics of wearable sensor data

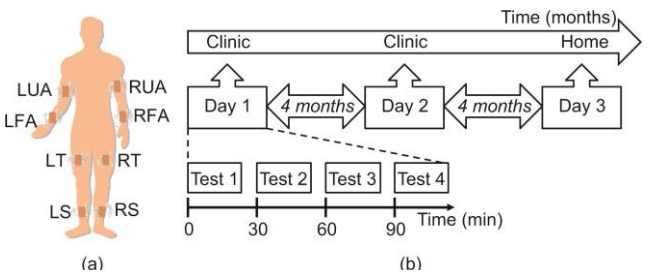

Figure 1 (a) The placement of the sensors on patients' body. (b) A graphical summary of the experimental procedures.

segments of 5 s duration and the clinical scores of LID based on 30 s visual observation of the subject's motor behavior.

II. Methods

A. Data Collection

Seven patients diagnosed with PD were recruited (6 males; 62.7±10 years). All patients were on levodopa therapy and three patients were on dopamine agonists.

The accelerometer data used in this study was collected over three visits that were scheduled approximately four months apart (Figure 1b). Each visit consisted of four test sessions that took place every 30 min. During each test session, subjects were asked to perform a series of motor tasks that consisted of items of the UPDRS as well as activities of daily living. Each motor task was performed during a trial. Each trial lasted 30 s. Video recordings of the subjects during each trial were used by two trained clinicians to generate scores of the severity of dyskinesia for each limb (right arm, left arm, right leg, and left leg) based on a scale resembling the Rush Dyskinesia Scale (0-Absent; 1-Minimal; 2-Moderate; 3-Intense; 4-Violent) [7]. The clinical scores reflected the most severe dyskinetic movements observed during the performance of each task over the 30 s trial. Subjects who participated in this study had dyskinesia scores ranging from 0 to 2. The following 12 tasks were performed during each session: finger to nose, finger tapping, alternating hand movements, leg agility, opening and closing of the hands (performed by each hand separately), sitting, and postural holding of the arms. Wearable accelerometers were placed on eight body segments (Figure 1a) (right and left upper arm (RUA and LUA), right and left forearm (RFA and LFA), right and left thigh (RT and LT), and right and left shank (RS and LS)). The wearable sensors used in this study were tri-axial accelerometer sensors based on the SHIMMER platform [8].

B. Overview of the Data Analytics

Figure 2a shows a schematic representation of the proposed algorithm to estimate the severity of LID from the sensor data. Accelerometer data for each limb was collected during the performance of the above-listed motor tasks separately for each trial. The data for a given limb was used for subsequent analyses only if that limb was not involved in the performance of the motor task under consideration. For example, the severity of LID for the right arm was estimated from all the motor tasks except the right finger-to-nose movements, right hand movements, and right alternate hand movements.

C. Feature Extraction

The raw accelerometer data was band-pass filtered with cutoff frequencies of 0.5 and 3.5 Hz using an IIR filter in order to remove noise and gross changes in the orientation of the body segments. Each task was divided into 30 epochs of 5 s.

Features that captured intensity/amplitude, smoothness and periodicity of the movement were extracted from each data segment. Intensity/amplitude of movement was represented by the root mean square value of the linearly de-trended accelerometer, velocity (integration of acceleration), and displacement (integration of velocity) time series. The smoothness of movement was represented by the standard deviation of the acceleration, velocity, and displacement time series as well as by the difference between the maximum and

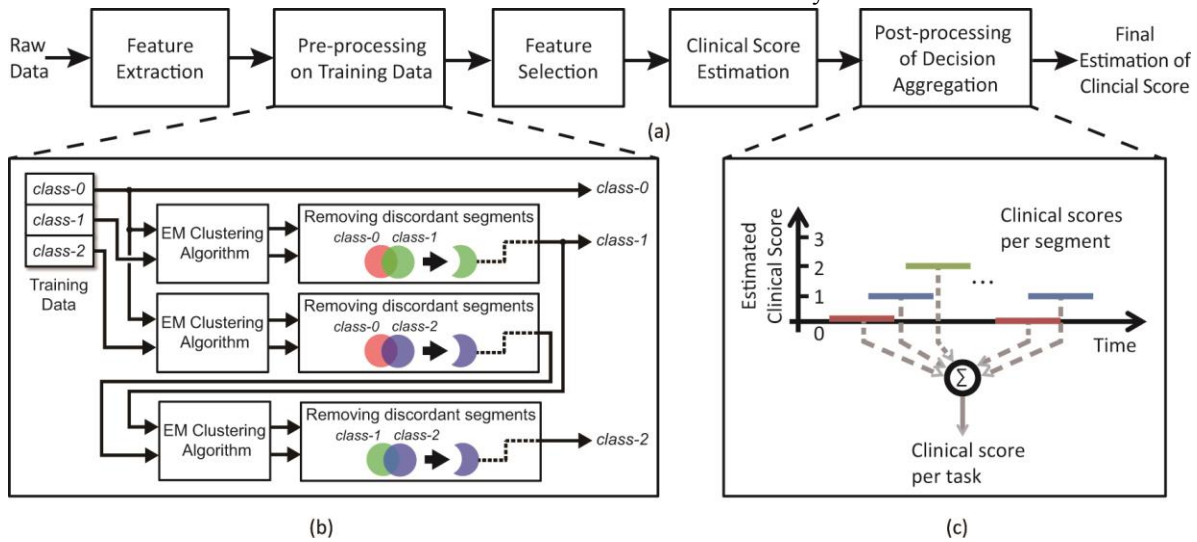

Figure 2 (a) Block diagram of the proposed algorithm to estimate the severity of LID. (b) A schematic representation of the pre-processing algorithm to remove data segments that appear to be mislabeled. (c) A schematic representation of the post-processing algorithm to aggregate the estimates derived from the performance of a single task to generate a single estimate.

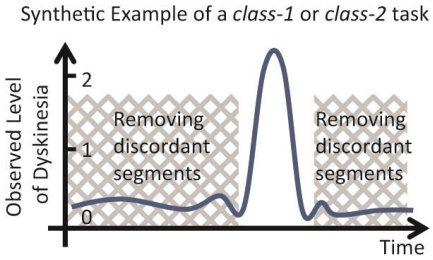

Figure 3 Graphical representation of the pre-processing algorithm to remove from the tasks data segments that are marked by a motor behavior that should be scored as 0 or 1.

the minimum values of the acceleration. The periodicity of movement was assessed by computing the dominant frequency and the ratio of the energy of the dominant frequency to the total energy. We also estimated the correlation coefficients between the accelerometer data collected using sensors on different limbs and the entropy of the accelerometer time series. These features (except for the inter-correlation between sensors) were computed from each axis of the accelerometer yielding 81 features per limb.

D. Pre-Processing Algorithm

Clinical scores were assigned to the performance of motor tasks based on the observation of the worst dyskinetic movements during each trial. Consider as an example the graphical representation shown in Figure 3 where involuntary movements are observed for a short period of time during the trial, which results in a label of score 2 for the whole trial. This implies that all the data segments pertaining to the trial, including those that show minimal involuntary movements, will be labeled as *class-2*. This will create mislabeled data segments, which compromises our ability to build a classification model.

The proposed pre-processing technique removes the mislabeled data segments. A schematic representation of the proposed pre-processing technique is shown in Figure 2b. The algorithm relies upon the expectation maximization (EM) clustering algorithm [9] to select data segments that were properly labeled. First, the *class-0* and *class-1* data of the training set were analyzed with the EM algorithm. The number of clusters was set to two apriori. Then, the data points of *class-1* that were assigned to the cluster of *class-0* were removed from the training set. Similarly, the EM algorithm was applied to *class-0* and *class-2,* and *class-1* and *class-2* in order to eliminate the datasets of *class-2* that show similar movement patterns as *class-0* and *class-1*.

E. Feature Selection

The cardinality of the feature set was reduced to remove any redundant information using the ReliefF algorithm [10] to rank the features and the Davies-Bouldin (DB) cluster validity index [11] to select the optimal cardinality. The ReliefF algorithm is an iterative algorithm that assigns weights to features based on their classification ability. This algorithm computes the weight of each feature by repeatedly sampling an instance and by investigating its *K* nearest neighbors of the same and different classes. Then, the features sorted according to their ranking were added progressively and the DB indices were computed. The DB index captures the intra- and inter-class separation of clusters. The feature cardinality that minimized the DB index was selected.

TABLE I The number of instances in each class for all limbs. This table shows that the dataset is unbalanced.

	class-0	*class-1*	*class-2*
Right Arm	572	15	0
Left Arm	535	48	4
Right Leg	646	137	53
Left Leg	688	110	37

F. Clinical Score Estimation

This work used a cost-sensitive classifier with the Random Forest (RF) algorithm as its base classifier in order to train the classification model and estimate the clinical scores. A cost-sensitive classifier is a meta-classifier that makes its base classifier cost sensitive by training the model to estimate a class while minimizing the misclassification cost defined by a cost matrix. A cost-sensitive classifier was employed in this work due to the unbalanced dataset, as the number of instances labeled as *class-0* was significantly higher than *class-1* and *class-2*. The cost matrix was defined to minimize the variance of root mean square errors (RMSE) of the estimations of the three classes. The RF algorithm is a boosted learning method based on a large collection of weakly-correlated trees of randomized features and data samples [9]. The RF algorithm is known to be robust to overfitting, which is an important property given the small sample size of the current study.

G. Post-processing Algorithm

Similar to the training set, the testing set may contain data segments that do not include movement patterns consistent with the clinical scores. We developed a post-processing algorithm to aggregate the estimates derived from data segments associated with each trial (Figure 2c). We tested four different decision aggregation techniques: 1) averaging the estimated values over the time interval corresponding to each trial, 2) taking a majority vote, 3) selecting the maximum estimated value, and 4) averaging of the estimated values using 20% of the data segments associated with the largest estimated scores. These aggregation techniques generated a single estimate per trial.

Note that we used the WEKA implementations of the EM, ReliefF, and RF algorithms for this study [12].

III. Results

Table I summarizes the number of instances associated with each class for each limb. The dataset is clearly unbalanced. Consequently, the RMSE of the clinical scores, which captures the difference between the scores provided by the clinical staff and the scores estimated on the basis of the sensor data, was computed separately for each class.

Figure 4 shows box-plots of the estimates achieved using the proposed algorithm for each class and for each limb. The

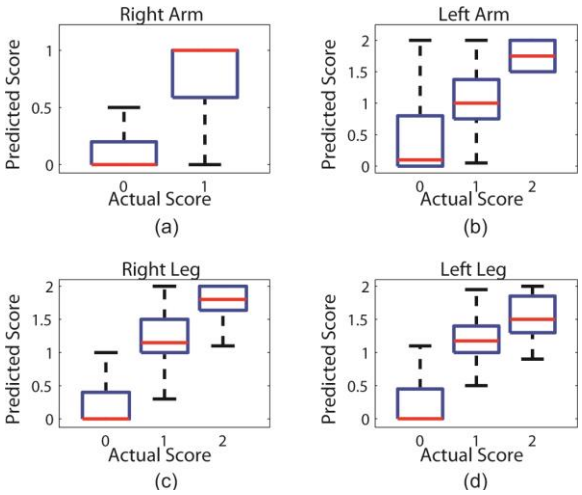

Figure 4 Bar-plots of the estimations made by the proposed algorithm for each class for (a) right arm, (b) left arm, (c) right leg, and (d) left leg.

TABLE II The average RSME values between the actual scores and the estimated scores, without processing and with both pre- and post-processing for each limb. The average RMSE values are computed for each class (in parentheses) and across all classes.

	RMSE	
	No Processing (*class-0, -1, -2*)	With Processing (*class-0, -1, -2*)
Right Arm	0.48 (0.46, 0.52, N/A)	0.44 (0.42, 0.45, N/A)
Left Arm	0.75 (0.59, 0.62, 1.02)	0.41 (0.67, 0.56, 0)
Right Leg	0.65 (0.57, 0.66, 0.64)	0.49 (0.47, 0.55, 0.43)
Left Leg	0.66 (0.68, 0.64, 0.66)	0.55 (0.50, 0.50, 0.64)

estimation errors were derived using a leave-one-subject-out cross validation (LOSOCV) technique. These results employed the averaging technique as the post-processing as it outperformed the other techniques. The average RMSE values were 0.44, 0.41, 0.49, and 0.55 for each limb, respectively.

The results in Table II show the average RMSE values between the actual scores and the estimated scores, without any processing and with both pre- and post-processing for each limb. The average RMSE values are computed for each class (in parentheses) and across all classes. Incorporating both pre- and post-processing showed the largest improvement of RMSE by 0.16, i.e. 0.043 for right arm, 0.34 for left arm, 0.14 for right leg, and 0.12 for left leg.

IV. Discussion and Conclusion

The results presented in this paper show that employing the proposed pre- and post-processing algorithms improves the accuracy of the estimated LID scores by 0.16 points (on a 0-2 point range spanned by the data used in the study). Applying only the pre-processing algorithm reduced the RMSE, on average, by 0.07 points. Applying only the post-processing algorithm reduced the RMSE by 0.05 points. The largest improvement in RMSE was observed for the left arm (Table II). In this specific case, the estimation algorithm without pre- and post-processing could not classify any of the *class-2* instances while the proposed techniques classified all *class-2* instances correctly. However, there were only four instances that were labeled as *class-2* for the left arm (Table I).

Our initial hypothesis was that selecting the maximum estimated LID score for a given trial or the average of the 20% largest estimated LID scores within each trial would have led to the most accurate results. However, the study showed that these techniques are too sensitive to outliers. Consequently the best results were achieved by averaging all LID score estimates derived for each trial.

This paper introduced novel pre- and post-processing techniques that reduce the effects of mislabeled data segments in the training dataset and effectively aggregate the estimation results derived from the testing dataset. The combination of the techniques led to accurate estimates of the severity of LID. The proposed method significantly improves the estimation of limb-specific clinical scores and can enable new research and clinical applications of wearable sensors with focus on LID.

References

[1] S. Fahn, "The spectrum of levodopa-induced dyskinesias," *Ann Neurol,* vol. 47, pp. S2-9; discussion S9-11, Apr 2000.

[2] B. Carignan, J. F. Daneault, and C. Duval, "Assessing drug-induced dyskinesia in the clinic, the laboratory and the natural environment of patients," *J Parkinsons Dis,* vol. 1, pp. 329-37, 2011.

[3] S. Patel, K. Lorincz, R. Hughes, N. Huggins, J. Growdon, D. Standaert, *et al.*, "Monitoring motor fluctuations in patients with Parkinson's disease using wearable sensors," *IEEE Trans Inf Technol Biomed,* vol. 13, pp. 864-73, Nov 2009.

[4] M. G. Tsipouras, A. T. Tzallas, G. Rigas, S. Tsouli, D. I. Fotiadis, and S. Konitsiotis, "An automated methodology for levodopa-induced dyskinesia: assessment based on gyroscope and accelerometer signals," *Artif Intell Med,* vol. 55, pp. 127-35, Jun 2012.

[5] B. T. Cole, P. Ozdemir, and S. H. Nawab, "Dynamic SVM detection of tremor and dyskinesia during unscripted and unconstrained activities," *Conf Proc IEEE Eng Med Biol Soc,* vol. 2012, pp. 4927-30, 2012.

[6] B. T. Cole, S. H. Roy, C. J. De Luca, and S. H. Nawab, "Dynamical learning and tracking of tremor and dyskinesia from wearable sensors," *IEEE Trans Neural Syst Rehabil Eng,* vol. 22, pp. 982-91, Sep 2014.

[7] C. G. Goetz, G. T. Stebbins, H. M. Shale, A. E. Lang, D. A. Chernik, T. A. Chmura, *et al.*, "Utility of an objective dyskinesia rating scale for Parkinson's disease: inter- and intrarater reliability assessment," *Mov Disord,* vol. 9, pp. 390-4, Jul 1994.

[8] (2015). *Shimmer Sensing Technology*. Available: http://www.shimmersensing.com/

[9] I. H. Witten, E. Frank, and M. A. Hall, "Data Mining: Practical Machine Learning Tools and Techniques (Third Edition)," I. H. W. F. A. Hall, Ed., ed Boston: Morgan Kaufmann, 2011.

[10] M. Robnik-Šikonja and I. Kononenko, "Theoretical and Empirical Analysis of ReliefF and RReliefF," *Machine Learning,* vol. 53, pp. 23-69, 2003/10/01 2003.

[11] J. C. Bezdek and N. R. Pal, "Some new indexes of cluster validity," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on,* vol. 28, pp. 301-315, 1998.

[12] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *SIGKDD Explor. Newsl.,* vol. 11, pp. 10-18, 2009.