Prediction of Parkinson's disease severity through MDS-UPDRS ratings and H&Y stage from IMU signals using neural network model

Ryu H, Wang G, Ph.D., Lee M, M.D., Nair R, Ph.D.

Division of Biostatistics, University of California, San Diego

UC San Diego

Abstract

Parkinson's disease (PD) is a progressive neurodegenerative disorder that accompanies motor symptoms. Prediction of PD symptoms from IMU sensor data in wearable devices is essential in monitoring PD progression. This project uses neural network model with LSTM and CNN layers (MLSTM-FCN) to predict PD symptoms with MDS-UPDRS II severity, MDS-UPDRS III severity, and H&Y stages as outcome variables. Model evaluation measures including accuracy, recall, precision, and F1 score are reported. The cross-validation accuracy is 0.9, 0.69, and 0.77 for MDS-UPDRS II, III, and H&Y respectively. This project aims to serve as an exploratory reference for future models in PD symptoms prediction using IMU sensor data and the neural network model.

Introduction

- Parkinson's disease (PD) is a progressive neurodegenerative disorder accompanying various motor and non-motor symptoms which includes freezing of gait (FoG), tremors, rigidity in body, sleep disturbances, amnesia, or mood change.
- The Movement Disorder Society-Sponsored Revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS), along with Hoehn and Yahr (H&Y) stages, is a comprehensive assessment of motor and non-motor symptoms associated with PD.
  - **MDS-UPDRS Part II** is a self evaluation of the activities of daily life.
  - **MDS-UPDRS Part III** is a motor evaluation monitored and scored by clinicians.
  - **H&Y** stages represents PD symptom progression with scales 1-5 administered by clinicians.
- The detection and monitoring of such symptoms at home proves particularly advantageous for PD patients, as home monitoring enables more constant and frequent monitoring without the inconvenience associated with physically visiting clinical facilities.
- Especially, wearable device such as smartphones or smartwatch equipped with Inertial measurement unit (IMU) sensor is essential in achieving objective PD progression monitoring.
- There exist a research gap in PD detection using IMU in that:
  1. Most existing literatures use kinetic features derived from IMU data[1] and not the raw IMU signals.
  2. Even those that use IMU raw signal only predict the PD status, and not the PD severity represented by clinical ratings.[2]
- To this end**, we are exploring the possibility of predicting the PD severity represented by MDS-UPDRS II**, **III, and H&Y stages, through IMU** consisting of 3 accelerometer and 3 gyroscope signals.
- Our project has following advantages:
  1. Using raw signals has benefits as the model does not have to depend on the proprietary feature extraction mechanism which is not always publicly available. Also, such difference in extraction mechanism makes comparison between models difficult.
  2. Using clinical ratings instead of PD status provides more granular prediction model which is essential for PD progression monitoring.
- Considering the time-dependent nature of the IMU signals, **we use the neural network model with long short-term memory (LSTM) and convolutional neural network (CNN)** layers.
- Multivariate LSTM-FCNs[3] model, which was developed for time-series analysis, is applied as a prediction model to a public data set that contains the IMU signals and clinical information of 35 PD patients with differing PD severity.

Data Description

- The data set is collected by Movement Disorders Clinic in the School of Medicine at the University of São Paulo.[4]
- The data set contains IMU signals and clinical information of **35** idiopathic PD patients, each with **up to three 120-second sessions**. The patients perform two turning motions within one session. Data is collected during **ON medication condition** with a stable dose of antiparkinsonian medication for at least 1 month, and dopaminergic medication one hour before each sessions.
- The clinical information includes total scores for MDS-UPDRS II, III and H&Y stages per session. The total score ratings of MDS-UPDRS are categorized into PD severity according to Martinez-Martin et al (2015)[5].
  - MDS-UPDRS II: 0-12 mild; 13-29 moderate; 30-52 severe
  - MDS-UPDRS III: 0-32 mild; 33-58 moderate; 59-132 severe
- MDS-UPDRS II and III severity, and H&Y rating per session (**total = 71**) are used as outcome variables with following distribution:
  - **H&Y**: "2"-10; "3"-57; "4"-4
  - **MDS-UPDRS II severity**: mild-64, moderate-7
  - **MDS-UPDRS III severity**: mild-46, moderate&severe-25

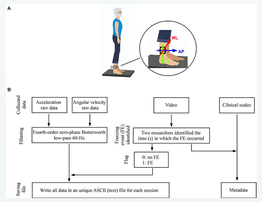


Figure 1: Figure explaining the placement of the IMU sensor (A) and the pre-processing procedure of the signals (B). "Acceleration raw data" comes from accelerometer, and "Angular velocity raw data" comes from gyroscope. IMU sensor *(Physilog 5 by Gait Up)* was placed on the leg of the most affected side at 128Hz (128 data points per second). IMU sensor outputs three accelerometer (mediolateral, anteroposterior, and vertical) and three gyroscope (mediolateral, anteroposterior, and vertical) signals. Figure source: Ribeiro De Souza et al (2022)[4]

Method

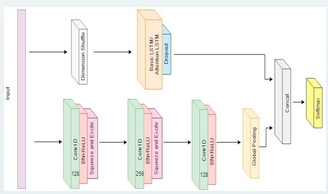


Figure 2: The schematic explains the multivariate long short term memory fully convolutional (MLSTM-FCNs) architecture. It has parallel structure where the input goes through a) LSTM layer and b) series of CNN layers. The two components are later concatenated before the output layer with softmax activation function. The model was trained with cross entropy loss for classification task. This project used basic LSTM layer. Figure source: Karim et al. (2019)[3]

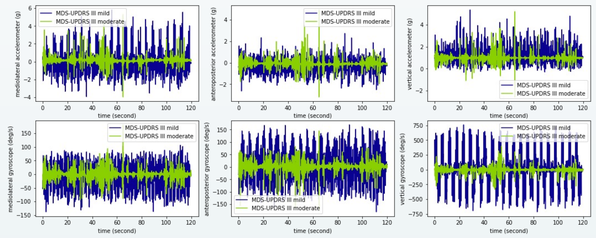
- To explore the viability of PD severity prediction model using neural network with IMU signals inputs, multivariate long short-term memory fully convolutional (MLSTM-FCNs)[3] model was applied.
- For each model, the input data and label were defined as the following:
  - Input data (X) : matrix of dimension *(n X (128*m) X 6)* where n is the number of windows, m is the window length, and 6 corresponds to 6 signals from IMU sensor. Each row represents a window with *128*m* columns, representing data points within each window.
  - Label (y): MDS-UPDRS-II severity ("mild" or "moderate"), MDS-UPDRS-III severity ("mild" or "moderate/severe"), and H&Y label ("2", "3", or "4")
- Subject-level k-fold cross-validation (k=5) was used to evaluate model performance with the following settings.
  - 60s window length with MLSTM-FCN architecture
  - 120s window length with MLSTM-FCN architecture with only CNN layers.

Results

Table1: Basic statistics (mean, standard deviation, minimum, and maximum) of the variables age, weight, height, MDS-UPDRS-II, MDS-UPDRS-III, H&Y stages, and total time in FoG per session.

variables	count	mean	std	min	max
Age	35	65.03	10.59	44	84
Weight (kg)	35	69.04	12.91	47	100
Height (m)	35	1.63	0.1	1.46	1.89
H&Y-Session1	35	2.91	0.51	2	4
H&Y-Session2	23	2.91	0.42	2	4
H&Y-Session3	13	2.92	0.28	2	3
MDS-UPDRS-II-Session1	35	8.54	3.78	1	21
MDS-UPDRS-II-Session2	23	7.17	3.27	2	18
MDS-UPDRS-II-Session3	13	7.46	2.67	5	14
MDS-UPDRS-III-Session1	35	31.83	14.99	9	78
MDS-UPDRS-III-Session2	23	28.26	13.41	11	68
MDS-UPDRS-III-Session3	13	24.62	12.02	11	52
Tot time in FoG (s)-Session1	35	27.22	39.77	0	120
Tot time in FoG (s)-Session2	23	20.87	38.8	0	116.3
Tot time in FoG (s)-Session2	13	13.8	32.96	0	120

Figure 3: The plots show exemplary accelerometer and gyroscope signals trajectories from a session. The blue represents the signal trajectories of patient 26 (session1) who is in MDS-UPDRS III mild group, whereas the green represents the signal trajectories of patient 2 (session1) who is in MDS-UPDRS III moderate group.



Model	H&Y 60s	H&Y 120s	Model	MDS-UPDRS-II 60s	MDS-UPDRS-II 120s	Model	MDS-UPDRS-III 60s	MDS-UPDRS-III 120s
Accuracy	0.771	0.785	Accuracy	0.900	0.901	Accuracy	0.685	0.671
Precision (2)	0.000	-	Precision (mild)	0.923	0.923	Precision (mild)	0.683	0.667
Precision (3)	0.781	0.785	Precision (moderate)	0.500	0.750	Precision (moderate/severe)	0.833	1.000
Recall (2)	0.000	0.000	Recall (mild)	0.972	0.973	Recall (mild)	0.978	1.000
Recall (3)	0.980	1.000	Recall (moderate)	0.250	0.438	Recall (moderate/severe)	0.113	0.040
F1 (2)	-	-	F1 (mild)	0.947	0.947	F1 (mild)	0.804	0.800
F1 (3)	0.869	0.880	F1 (moderate)	0.333	0.553	F1 (moderate/severe)	0.199	0.077
F1	0.869	0.880	F1	0.640	0.750	F1	0.502	0.439

Table 2: The table shows the model accuracy, precision, recall and F1 score per class, and average F1 score after 5-fold cross validation. For each fold, the session-level confusion matrix was first computed by aggregating the window-level predictions by taking the majority. Then, session-level confusion matrix was based on the session-level prediction and true labels. For the three outcome variables H&Y, MDS-UPDRS II and III, window length 60s and 120s were used. The "-" stands for the unavailable output, which corresponds to the case where the denominator was 0 (no prediction or true label for that class). Recall(4) was 0 for both H&Y models (omitted from table for space). Precision(4) and F1(4) for H&Y are omitted due to unavailable output.

Confusion Matrix		Predicted label 0	Predicted label 1	
True label	0	A	B	$\text{precision}_0 = \frac{A}{A+B}$
True label	1	C	D	$\text{precision}_1 = \frac{D}{C+D}$
		$\text{recall}_0 = \frac{A}{A+C}$	$\text{recall}_1 = \frac{D}{B+D}$	$\text{Accuracy} = \frac{A+D}{A+B+C+D}$

- Class-specific F1 score is defined as the harmonic mean of recall and precision, namely:
  - $F1_c = \frac{2 * \text{recall}_c * \text{precision}_c}{\text{recall}_c + \text{precision}_c}$
- Overall F1 score is defined as the unweighted average of class-specific F1 scores.

Conclusion

- This project explored the use of raw IMU signals in PD symptom severity prediction.
- Application of neural network models with LSTM and CNN layers to IMU signals made of different window lengths was proposed with outcome variable as H&Y and PD severity based on MDS-UPDRS II and MDS-UPDRS III ratings.
- The six different models, of different outcome variables and window length, were compared in term of accuracy, precision, recall, and F1 score which were calculated from 5-fold cross validation on based on session-level predictions.
- **H&Y** rating prediction showed cross validation accuracy of 0.77 and 0.79 for the 60s and 120s models.
- **MDS-UPDRS II** severity prediction showed cross validation accuracy of 0.9 for the 60s and 120s models.
- **MDS-UPDRS III** severity prediction showed cross validation accuracy of 0.69 and 0.67 for the 60s and 120s models.
- Due to the imbalance among classes present across the outcome variables, the classes with lower number of data points showed lower, if not unavailable, recall and precision, namely classes 2 and 4 for H&Y, moderate for MDS-UPDRS II, and moderate/severe for MDS-UPDRS III.
- The F1 scores, which is the harmonic mean of recall and precision, were 0.87 and 0.88 for H&Y, 0.64 and 0.75 for MDS-UPDRS II, and 0.50 and 0.44 for MDS-UPDRS III for 60s and 120s models.

Future Works

- This project's contribution stands at the fact that, to the authors' knowledge, there are no existing attempts of predicting clinical ratings from IMU raw signals
- However, with the presence of unbalanced outcome classes, the class-level prediction was unstable. The results might not be generalizable to patients with moderate or severe PD severity as they are under-represented in the analysis.
- Regarding this, correctional measures such as using weighted loss function can be considered.
- Other future works also include the following:
  - Incorporating the feature data or patient demographic/clinical information to the model to strengthen prediction power.
  - Incorporating multiple sensors measured from different locations of the body.
  - Incorporating IMU data that include more than one activity from patients with various degrees of PD symptoms.
  - Applying different types of prediction task such as models supposing outcome variables as count (clinical ratings as discrete positive values) or continuous (clinical ratings as continuous values) data.

Acknowledgements

This work was funded by the Oak Ridge Institute for Science and Education (ORISE) fellowship program.

1. Sotirakis, Charalampos, et al. "Identification of motor progression in Parkinson's disease using wearable sensors and machine learning." npj Parkinson's Disease 9.1 (2023): 142.
2. Lin, Chin-Hsien, et al. "Early detection of Parkinson's disease by neural network models." IEEE Access 10 (2022): 19033-19044.
3. Karim, Fazle, et al. "Multivariate LSTM-FCNs for time series classification." Neural networks 116 (2019): 237-245.
4. Ribeiro De Souza, Caroline, et al. "A public data set of videos, inertial measurement unit, and clinical scales of freezing of gait in individuals with parkinson's disease during a turning-in-place task." Frontiers in Neuroscience 16 (2022): 832463.
5. Martínez-Martín, Pablo, et al. "Parkinson's disease severity levels and MDS-Unified Parkinson's Disease Rating Scale." Parkinsonism & related disorders 21.1 (2015): 50-54.