

# Deep Learning Pose Estimation model for Parkinsonism and Levodopa-Induced Dyskinesia

Yashnil Saha

Monta Vista High School, Cupertino, CA, USA, 95014

**KEYWORDS.** Parkinson's disease, Levodopa-induced dyskinesia, Unified Parkinson's Disease Rating Scale, Unified Dyskinesia Rating Scale

**BRIEFS.** This research project utilized a Deep Learning Classification Model to design an algorithm for the detection of pathological motion, whether Parkinson's or dyskinesia, and the prediction of the clinical rating of Parkinson's or dyskinesia severity based on movement features.

**ABSTRACT.** Diagnosing Parkinson's disease (PD) is one of the largest challenges healthcare systems face due to the absence of a specific test for the condition and symptoms varying widely from person to person. Designing an automated model to aid in early diagnosis would greatly contribute to solving this problem. Currently, diagnosis for PD relies on clinical evaluation which has an error rate of approximately 20% [1], indicating the urgent need for an automated system to be developed. Levodopa is used for the treatment of PD but can lead to motor complications known as levodopa-induced dyskinesia (LID) when taken for too long. PD and LID are evaluated according to the Unified Parkinson's Disease Rating Scale (UPDRS) and Unified Dyskinesia Rating Scale (UDysRS) scales, respectively, which range from 0 to 4 (0-normal, 4-severely impaired) [2,3]. The tests are conducted by medical personnel and are very subjective because they rely on the experience of the rater. The goal of this project was to design an algorithm using deep learning for assessment of PD and LID using pose estimation. Two models were created: a regression model to predict the clinical rating from 0 to 4 and a classification model to determine whether the patient had PD or LID. During the feature extraction process, 32 features were extracted per joint trajectory including 15 kinematic, 16 spectral, and the convex hull of the movements. Then, the two neural network models were trained on these features to be able to predict their respective targets. The classification model achieved a mean F1-score greater than 0.86 and the regression model attained a root mean square error less than 0.49 for the Communication task, proving that this project was a promising start in the venture to automate diagnosis of PD.

## INTRODUCTION.

Nearly one million people in the United States live with PD and an array of associated disorders collectively known as parkinsonism [4]. At the present, diagnosis for PD relies on clinical evaluation which has an error rate of approximately 20% [1]. PD is a central nervous system disorder that affects physical movement, causing symptoms such as tremors, slowness, stiffness, and more physical symptoms. The cause of these involuntary movements begins in the brain, where the production of dopamine, the neurotransmitter that controls movement, is impaired. Without enough dopamine, the symptoms of PD become more severe. The main types of movement disorders that people with Parkinson's may experience are tremors, bradykinesia, rigidity, dyskinesia, dystonia, freezing, drooling and gait disorder [5]. Since the discovery of levodopa in 1960, it has been used for the treatment of PD and is powerful enough to improve motor symptoms [6]. Extended use of levodopa has been found to cause levodopa-induced dyskinesia (LID) within 4–6 years in 40% of individuals [7]. LID refers to involuntary adventitious movements that usually occur after prolonged treatment with levodopa in PD patients. The term dyskinesia is applied to any involuntary movement, such as chorea, ballism, dystonia, tic, or myoclonus. The most common types of levodopa-induced dyskinesia are chorea and dystonia, which often coexist. Myoclonus, ballism, tics, or stereotypy are far less common

[8]. Although PD patients follow up consistently with their neurologists for consultation, these follow-ups are not regular causing significant changes in a patient which are difficult to determine. Also, the clinical rating scales which are used to measure the PD symptoms require special trained personnel and are very subjective. This depends a lot on the skill of the nurses and clinician staff [9]. Sometimes patients use paper diaries for recording their symptoms, however making sense of the symptoms varies between patients and the doctor [10,11]. Clinical diagnosis completed per the Unified Parkinson's Disease Rating Scale (UPDRS) [2,12] and Unified Dyskinesia Rating Scale (UDysRS) [3] is used for assessing the severity of a patient's disease.

Deep learning is a class of machine learning algorithms that uses multiple layers to progressively extract higher-level features from the raw input data. Deep learning has the capability to create models that are trained on a dataset and can be used for classification or regression as per the requirements of the problem domain. Deep learning has immense potential and can be used to solve problems in the medical field for diseases as it is more efficient and provides faster predictions. However, it suffers from a need for more computational resources and is constrained by the data it is trained on [13]. Pose estimation is a computer vision technique which is used to predict and track the location of a person or object. This is done by blending the pose and the orientation of a given person/object. Pose estimation has been used for recognizing motion, in Kinect [14], Convolutional Pose Machines [15] and Open Pose [16]. This process functions by recognizing the joints in the human body and using these to represent body parts. Using the joint information, it can then identify the gestures and actions performed by the human body. This leads to the measurement of movement from changes in posture of the individual undergoing examination. More precisely, the work done by Li et al [17] used a Vision-based assessment of PD and levodopa-induced dyskinesia with pose estimation. Video input of patients performing some particular activities was used as the input data. Using the Convolutional Pose Machines, the distinctive movement of the joints were extracted to build a pose estimation model to classify and predict the scores of UPDRS and UdysRS. Sato et al. [18] incorporated a method which evaluated the disease by figuring out the length and duration of the gait and rhythm. The work utilized videos of healthy individuals and patients with Parkinson's. They used the OpenPose technique and an unsupervised machine learning model.

## MATERIALS AND METHODS.

### *Neural Networks.*

Neural networks were utilized for the model through TensorFlow, a set of tools and libraries for machine learning [19]. The core building block of a neural network is a layer which helps in extracting representations out of the data fed into them. Deep learning consists of chaining together simple layers to implement a form of progressive data distillation [20]. The algorithm finds a set of values for the model's weights that minimizes a loss function for a given set of training data samples and their corresponding targets. The entire learning process is made possible by the fact that all tensor operations

in neural networks are differentiable. Thus, it's possible to apply the chain rule of derivation to find the gradient function mapping the current parameters and batch of data to a gradient value. This process is known as backpropagation. While configuring the learning, a loss function is minimized during training. The optimizer determines how the network should be updated based on the loss function. Metrics are a measure of success during training and validation.

#### Data.

The model used in this research utilized the identical dataset from previous work by M.H. Li, T. A. Mestre, S. H. Fox, and B. Taati. [17]. The acquired Kaggle dataset [21] contained pose estimates. The data was produced at the Center for Movement Disorders of Toronto Western Hospital. There were nine participants with PD and LID, 5 men and 4 women, with an average age of 64 years. The dataset was divided into separate files for PD and LID, allowing the model to predict each condition separately. These participants were given certain tasks to be performed according to the standard assessment scales of UPDRS and UDysRS. The UDysRS Part III was used to rate the severity of dyskinesia and the UPDRS Part III was used to rate the severity of parkinsonism. The actions of the participants were taken through a 480×640 or 540×960 video camera of 30 frames per second as they stood directly opposite, while evaluations were made at a frequency of 15-30 minutes for a period of 2-4 hours by three specialist neurologists. The tasks assigned to the participants were the following below:

- Communication – The participants were asked to describe an image, talk to the examiner, and answer mental math. This was conducted as per UdysRS Part III.
- Drinking – The participants were asked to drink from a cup. This was rated as per UdysRS Part III.
- Leg Agility – The participants were asked to stomp their legs with as much speed and amplitude as possible. This was rated as per UPDRS Part 3.8.
- Toe Tapping – This was rated as per UPDRS Part 3.7.

The features were extracted from the data. Convolutional Pose Machines (CPM) were used to find the joints of the human body from the videos. In this process 2D movement trajectories were extracted for head, neck, shoulders, elbows, wrists, hips, knees, and ankles. This study focused on building a regressor to predict the severity of PD and LID and a classifier to identify whether patients had PD or LID.

#### Pre-processing of Data.

In this stage of data processing, score thresholds were used to balance classes. For the communication and drinking tasks, a threshold of 0.5 was used for binarizing scores. For leg agility a threshold of 1 was used and for toe-tapping binarization 2 was used.

#### Feature Extraction from Movement Trajectories.

To improve the generalization potential of the model, a total of 32 separate features were extracted per joint trajectory except for toe tapping: 15 kinematic, 16 spectral, and the convex hull of the movements. The 15 kinematic features were namely the maximum, median, mean, standard deviation and interquartile range of the speed, magnitude of acceleration and magnitude of jerk. This study focused only on the scalar kinematic features or features with only magnitude as the magnitude of movement was more important than the direction. The 16 spectral features were computed from the Welch power spectral density (PSD) and included the peak magnitude, entropy, total power, half point (i.e., frequency that divides spectral power into equal halves), and power bands 0.5–1 Hz, > 2 Hz, > 4 Hz, > 6 Hz for both

the displacement and velocity PSDs. The final feature was the convex hull, which quantifies the area that a joint moved within.

#### Background and Dependencies.

The communication and drinking tasks were used to predict their respective UdysRS Part III item scores, while the leg agility and toe tapping tasks were used to predict their UPDRS Part III item scores. For each of the sub scores of the UdysRS and UPDRS, ratings were on a scale of 0–4, where 0 indicated normal motion and 4 indicated severe impairment. In order to build a deep learning model, Keras was used. Numpy, json, matplotlib, pandas, scipy, and seaborn were also leveraged. The regression model was chosen to predict the value from 0 to 4 that corresponded with the UPDRS and UDysRS scales as its output was a continuous value. The classification model was also included in order to be able to simply diagnose whether a person had either condition.

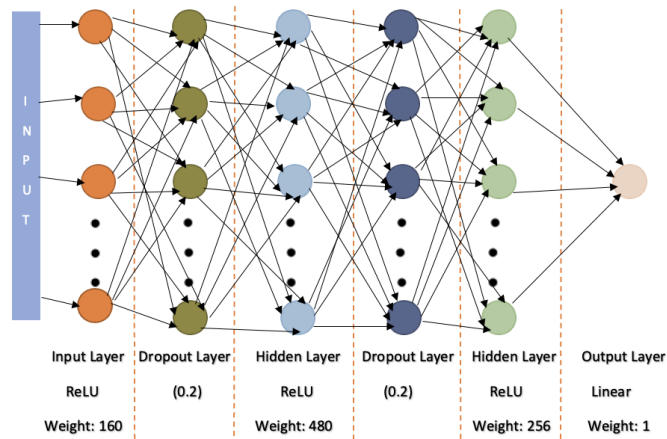
#### Regression Model.

To determine the clinical rating of PD or LID severity based on movement features, regression was used. The neural network model for the regressor is shown in Figure 1 which uses 3 dense hidden layers with 480, 256 and 1 neurons.

The hyperparameters for the regression model are summarized in Table 1.

#### Classification Model.

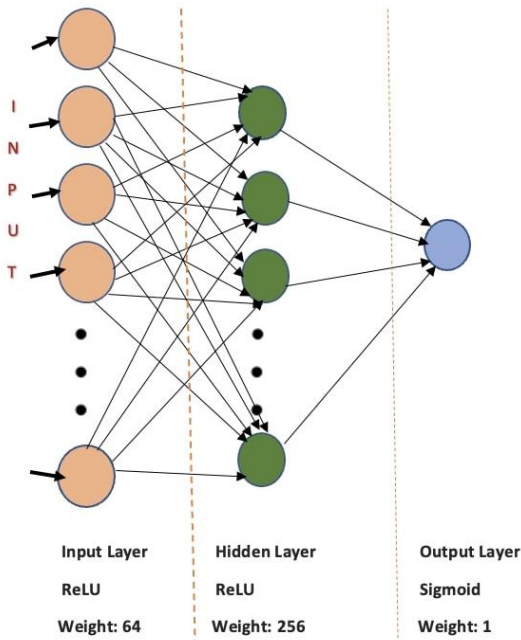
In contrast to the regression model which contained 3 dense hidden layers, the classification model contained only 2 due to it having a less complicated task. Dropouts were also not included in the classification model as they were found to be unnecessary in the simpler model. Binary classification was used to determine the pathological motion, whether the patient had PD or whether they had LID. The neural network model for the classifier is shown in Figure 2 and used 2 dense hidden layers with 256 neurons and 1 neuron.



**Figure 1.** Regression model diagram depicting layers of the neural network with corresponding weights and activation functions. There are dropout layers in between the first three layers with a 20% dropout probability.

**Table 1.** Hyperparameters for Regression Model.

Optimizer	Adam
Activation Function	ReLU, Linear
Loss Function	Mean Squared Logarithmic Error
Batch Size	64
Epochs	25
Number of dense hidden layers with respective number of neurons	3 with 480, 256, 1



**Figure 2.** Classification model diagram displaying three-layer neural network with corresponding weights and activation functions.

**Table 2.** Hyperparameters for Classification Model.

Optimizer	Adam
Activation Function	ReLU, Sigmoid
Loss Function	Binary Crossentropy
Batch Size	128
Epochs	20
Number of dense hidden layers with respective number of neurons	2 with 256, 1

The hyperparameters for the classification model are summarized in table 2.

## RESULTS.

The following tables display the results of the various models for each of the four tasks and individual joint trajectories. The metrics of F1-score and Area under the Curve were used to evaluate the Classification model and the metric of Root Mean Square Error was used to evaluate the Regression model. The Pearson coefficient was also included for the Regression model as an extra factor to help determine the correlation between the predicted values and actual values. Tables 7 and 8 showcase the superior performance of the model for the communication task in comparison to the other tasks, with a mean F1-score of 0.865 and Pearson coefficient of 0.709. The toe tapping and leg agility tasks performed similarly, with F1-scores of 0.861 and 0.831, respectively as seen in tables 3 and 5. However, the drinking task was by far the worst, with a mean F1-score of just 0.495 as displayed in table 9. It also had the lowest Pearson coefficient of just 0.171 as shown in table 10.

## DISCUSSION.

Overall, the results of the two models are very promising and prove the viability of this approach for diagnosing Parkinson’s disease and levodopa-induced dyskinesia. The communication task brings out involuntary movements and thus had the best performance. Drinking

**Table 3.** Classification performance metrics for Toe Tapping task.

	Left	Right	Mean
F1-score	0.825	0.896	0.861
AUC	0.500	0.642	0.571

**Table 4.** Regression performance metrics for Toe Tapping task.

	Left	Right	Mean
RMSE	0.550	0.541	0.545
r	0.206	0.485	0.345

**Table 5.** Classification performance metrics for Leg Agility task.

	Left	Right	Mean
F1-score	0.837	0.825	0.831
AUC	0.644	0.500	0.572

**Table 6.** Regression performance metrics for Leg Agility task.

	Left	Right	Mean
RMSE	0.365	0.462	0.414
r	0.593	0.243	0.418

**Table 7.** Classification performance metrics for Communication task.

	Neck	Rarm	Larm	Trunk	Rleg	Lleg	Mean
F1	0.908	0.860	0.860	0.910	0.820	0.830	0.865
AUC	0.925	0.874	0.878	0.915	0.739	0.804	0.856

**Table 8.** Regression performance metrics for Communication task.

	Neck	Rarm	Larm	Trunk	Rleg	Lleg	Mean
RMSE	0.505	0.389	0.467	0.617	0.341	0.585	0.484
r	0.784	0.523	0.743	0.864	0.537	0.804	0.709

**Table 9.** Classification performance metrics for Drinking task.

	Neck	Rarm	Larm	Trunk	Rleg	Lleg	Mean
F1	0.759	0.332	0.439	0.598	0.500	0.539	0.495
AUC	0.728	0.508	0.578	0.600	0.475	0.608	0.582

**Table 10.** Regression performance metrics for Drinking task.

	Neck	Rarm	Larm	Trunk	Rleg	Lleg	Mean
RMSE	0.473	0.531	0.712	0.509	0.511	0.560	0.549
r	0.234	0.172	0.173	0.087	0.120	0.241	0.171

task arm subscore performance was inferior to other subscores due to the inability to differentiate voluntary from involuntary movements and increased occlusion of upper limbs during movement. The communication task achieved a mean RMSE of 0.484 and Pearson coefficient of 0.709 among all joint trajectories. This validated that it was able to recreate a clinical evaluation very well. Although the RMSE was similar for the communication and drinking tasks, the Pearson coefficient of 0.171 was lower in the drinking task, showing

how its performance was comparatively worse. Most ratings for the drinking task were in the narrow range of 0 to 2, indicating the need for both RMS and Pearson coefficient to evaluate model performance. The communication task was the strongest for both classifying LID and determining its severity. For PD, leg agility was stronger for regression and toe tapping was stronger for classification.

## CONCLUSION.

The results of the deep learning model prove the viability of using Artificial Neural Networks to predict and classify PD and LID. Levodopa is the most effective Parkinson's disease medicine, however there are other treatments such as dopamine agonists which have their own side effects. Furthermore, the model currently only requires the computational power of a laptop and could be scaled up to work with a larger dataset. A more accurate dataset could be obtained directly from multiple health care providers to try and minimize the error that is inherently present within the dataset. In the future, an automated system could be developed to detect changes in the severity of symptoms to trigger clinical trials for new therapies.

## REFERENCES

1. Statistics on Parkinson's – Parkinson's Disease Foundation. <<https://www.parkinson.org/understanding-parkinsons/statistics>>
2. C. G. Goetz et al., "MDS-UPDRS-Revision of the Unified Parkinson's Disease Rating Scale", *Movement Disorders*. **23**, 2129-2170 (2008).
3. C. G. Goetz, J. G. Nutt and G. T. Stebbins, Unified Dyskinesia Rating Scale (UdysRS) *Movement Disorders*. **23**, 2398-2403 (2008).
4. C. W. Hess, M. S. Okun. *Diagnosing Parkinson Disease*. Continuum (Minneapolis). **22**, 1047-1063 (2016).
5. B. Upham. 8 Ways Parkinson's Disease Affects Your Movement <<https://www.everydayhealth.com/hs/parkinsons-disease/parkinsons-movement-types/>>
6. National Collaborating Centre for Chronic Conditions (UK). *Parkinson's Disease: National Clinical Guideline for Diagnosis and Management in Primary and Secondary Care*. London: Royal College of Physicians (UK); 2006 [cited 2015 Nov 28]. <<http://www.ncbi.nlm.nih.gov/books/NBK48513/>>
7. Ahlskog JE, Muenter MD. Frequency of levodopa-related dyskinesias and motor fluctuations as estimated from the cumulative literature. *Mov Disord*. **16**, 448–58 (2001).
8. Chong S. Lee, MD, FRCPC. Levodopa-induced dyskinesia: Mechanisms and management. *Issue: BCMJ*. **43**, 206-209 (2001).
9. Post B, Merkus MP, de Bie RMA, de Haan RJ, Speelman JD. Unified Parkinson's disease rating scale motor examination: are ratings of nurses, residents in neurology, and movement disorders specialists interchangeable? *Mov Disord*. **20**, 1577–84 (2005).
10. Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient compliance with paper and electronic diaries. *Control Clin Trials*. **24**, 182–99 (2003).
11. Goetz CG, Leurgans S, Hinson VK, Blasucci LM, Zimmerman J, Fan W, et al. Evaluating Parkinson's disease patients at home: utility of self-videotaping for objective motor, dyskinesia, and ON-OFF assessments. *Mov Disord*. **23**, 1479–82 (2008).
12. S. Fahn and R. Elton, *Unified Parkinson's disease Rating Scale – UPDRS*, 1987.
13. M. Javaid, A. Haleem, R. P. Singh, R. Suman and S. Rab, "Significance of machine learning in healthcare: Features pillars and applications", *International Journal of Intelligent Networks*. **3**, 58-73 (2022).
14. R. Miles, *Start here! learn the Kinect API*, Pearson Education, 2012.
15. S. E. Wei, V. Ramakrishna, T. Kanade and Y. Sheikh, "Convolutional Pose Machines", *CVPR*, 2016.
16. Z. Cao, G. Hidalgo, T. Simon, S. E. Wei and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields", *CVPR*, 2019.
17. M. H. Li, T. A. Mestre, S. H. Fox and B. Taati, "Vision-based assessment of parkinsonism and levodopa-induced dyskinesia with pose estimation", *Journal of NeuroEngineering and Rehabilitation*. **15**, 13-17 (2018).
18. K. Sato, Y. Nagashima, T. Mano, A. Iwata and T. Toda, "Quantifying normal and parkinsonian gait features from home movies: Practical application of a deep learning-based 2D pose estimator", *PLOS ONE*. **14**, 22-24 (2019).
19. TensorFlow <<https://github.com/tensorflow/tensorflow>>.
20. Francois Chollet. *Deep Learning with Python*, Second Edition. Published by Manning Publications.
21. Parkinson's Vision-Based Pose Estimation Dataset <https://www.kaggle.com/datasets/limi44/parkinsons-visionbased-pose-estimation-dataset>



Yashnil Saha is a student at Monta Vista High School in Cupertino, CA.