# Predictive Models for Early Diagnosis of Parkinson's Disease in Men

**Bashayr Alshammari[1] Dr. Laura[2]**
[1,2]Department of Computer Science
[1,2]University of Findlay, Ohio, US

*Abstract*— The recent developments in data analysis and artificial intelligence have been increasingly employed in various fields to solve our day-to-day problems. These techniques have been used in the field of medicine to ensure a high level of accuracy in the diagnosis of diseases. Parkinson's disease (PD) is among the conditions that can be accurately diagnosed using these methods. PD is a neurodegenerative illness common among the elderly in the United States- it is more rampant in older men than in women. There are several symptoms of this disease that can be used for diagnosing this disease in the early stages. However, many people neglect the initial signs as a consequence of age. Moreover, these symptoms are often mistaken with other diseases, thereby resulting in a delayed cure. This research aims to build three models for early detection of Parkinson's disease by comparing symptoms in diagnosed patients against unaffected individuals in order to predict whether a person will get tremors as they become older. Furthermore, it also analyzes the duration of the hold time, the flight time, and variations in the press duration to predict if a person will get Parkinson's disease. Finally, it aims to predict the consequences of PD based on tremors, hand, and directions as the disease progresses over time. This research and data analysis seek to assist the community and the medical field in general to discovering symptoms of a silent epidemic as early as possible to guarantee early treatment.

***Keywords:*** Parkinson's disease (PD)

## I. Introduction

Parkinson's disease (PD) is a neurodegenerative illness common among the elderly in the U.S. Data from the population on 36 million beneficiaries of Medicare over 65 years of age indicate that 1.6% of the American population are undertaking PD treatment yearly.

Cardinal symptoms of PD include shuffling gait, stiffness, limb tremor, slowness, postural instability, voice tremor, loss of smell, cramped handwriting, etc. Apart from the high risk of physical disability, PD is associated with the rising cases of deaths among the elderly.

## II. Operationalized Research Questions:

The following are some of the operationalized research questions in this regard:

– Do age and Parkinson's predict tremors as a second symptom? (Model-1)
– What features of Keystroke task can be meaningfully incorporated into a predictive model of PD? (Model-2 and Model-3)
– Is there relationship between PD and length of disease?(Correlation)

## III. Significance of the Research:

Symptoms of Parkinson's disease may be caused by many other diseases which may be misdiagnosed as PD. On the other hand, many people neglect the early symptoms as a consequence of age. Delayed diagnosis and the subsequent treatment for PD is less effective. Improvement in health after initiating treatment is another characteristic of Parkinson's disease. Moreover, early diagnosis is very important. The motivation for this project is Building three models by comparing Parkinson's disease patients with people without PD:

a) Predict if a person will get tremors as they become older.
b) Predict if a person will get PD.
c) Predict the consequences of PD.

## IV. Tools and Methods

– R studio for gathering data.
– SQL used to clean and filter the datasets as well as RStudio.
– Excel to clean dataset.
– SPSS to build regressions and analysis.
– Tableau for visualization.

## V. Data Collection and Processing:

The dataset was obtained from Physio Net site. This website is supported by the National Institute of General Medical Science (NIGMS) and the National Institute of Biomedical Imaging and Bioengineering (NIBIB). This data was obtained from 200 subjects. Each patient was made to press certain keys and the recordings of the press time duration for a period of one month were made. First, I cleaned and filtered the dataset based on the male gender and age between 75-85 years in the beginning. There were no missing or incomplete values in the data as it was collected by proper experimentation and this task took me one week to finish it. Next, two new columns were added, namely, Ages and Length Disease, as I thought that these variables might be significant. When I wanted to calculate the duration of disease, I found that some users didn't provide meaningful information, for example, some just put their birth year, etc. So I filtered out meaningless information, to get a more meaningful interpretation. It took me another week to complete this task and to make sure that the final data was clean-the final table, including 19 columns and 169,282 rows. Subsequently, I start using tableau for visualization. The reason for this choice is that I wanted to understand the relationship between the variables before doing the regression analysis. After that, I was performing logistic regression using the SPSS tool and Excel. I start my analysis by checking for multicollinearity between the predictors. Since length of the disease was set as null for normal people, it was falsely showing as an insignificant factor, so I replaced null with 0. After this we checked for a correlation between length of disease and every single independent variable to use this variable in my predictive model. The length of disease is significantly correlated with tremors $r(97276)=0.19$, $p < .01$,(two-tailed).

 

*1) Predict if a person will get tremors, based on age and PD:*

Without the independent variables, the model was 58.7% accurate. The prediction power improved to approximately 77.8% after the addition of the independent variables, as shown in Table1. From the pseudo R2 tests, it was found that the independent variables account for 48% to 65% of the variability of being affected with PD. All these variables were statistically significant with $p < 0.05$.

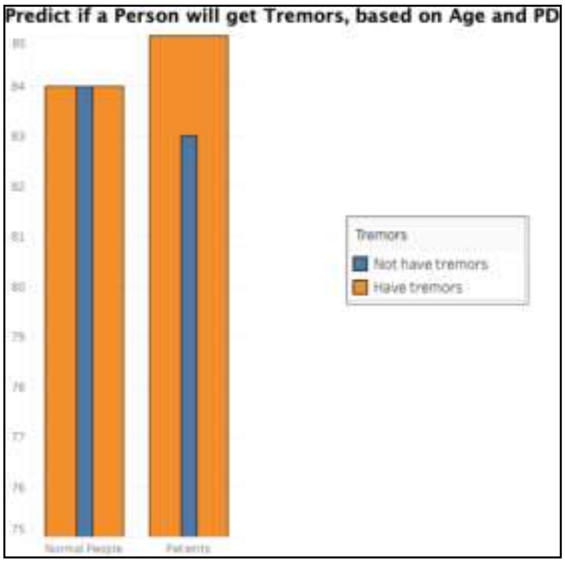


Fig. 1: First Model

## VI. RESULTS INTERPRETATION:

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I.for EXP(B) Lower | Upper |
|---|---|---|---|---|---|---|---|---|---|
| Step 1[a] | Ages | -.194 | .003 | 4927.791 | 1 | .000 | .824 | .819 | .828 |
| | Parkinsons(1) | 6.650 | .046 | 20722.588 | 1 | .000 | 772.630 | 705.751 | 845.848 |
| | Constant | 9.710 | .212 | 2090.624 | 1 | .000 | 16474.337 | | |

a. Variable(s) entered on step 1: Ages, Parkinsons.

Table 2. Logistic Regression Model Result

*1) Age has a negative coefficient:*

The odds of not being affected with tremors is 1.21 times more for every year increase in age ($b=-.194$, $p < .01$, $OR=.824$, (95%CI: .819, .828)).

*2) Individuals who have Parkinson's as second symptoms are 772.63 times more likely to have tremors than those who do not have Parkinson's ($b=6.650$, $p < .01$, $OR=772.63$, (95%CI: 705.75, 845.85)).*

*A. Predict if a person will get PD, based on age and hold time, flight time, and latency time:*

Before we start explaining the model, we would like to define times terms, As shown in table 3.

| | |
|---|---|
| Hold Time | A measure of how quickly the finger is tapped and can also indicate the relative force of tapping |
| Latency time | The time between successive keystrokes can be separated into same-hand(LL and RR) and opposite-hand(LR and RL). space key treated separately since they can use either hand to press on space key(LS, RS, SL, SR). |
| Flight Time | The elapsed time between releasing a key and pressing the subsequent key. |

Table 3: times terms

Without the independent variables, the model was 58.7% accurate. The prediction power improved to approximately 94.3% after the addition of the independent variables, as shown in Table1. From the pseudo R2 tests, it was found that the independent variables account for 48% to 65% of the variability of being affected with PD. All these variables were statistically significant with $p < 0.05$, except for the latency time, which I removed from the model

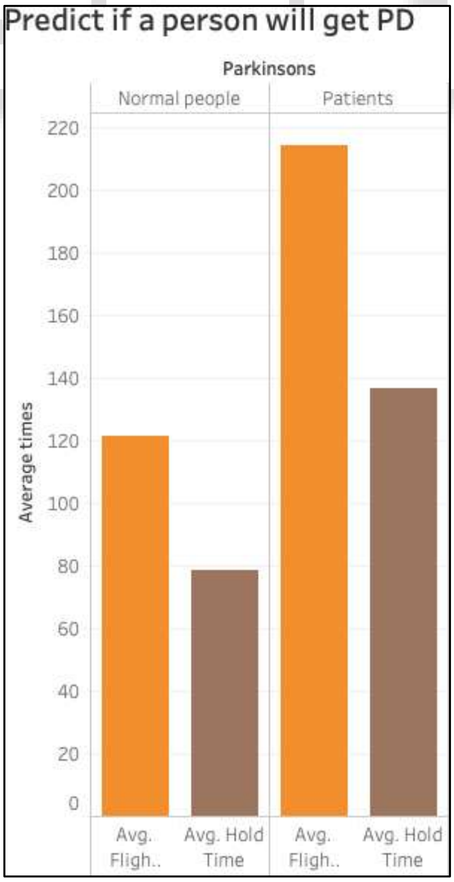


Fig. 2: Second Model

## VII. RESULTS INTERPRETATION:

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I.for EXP(B) Lower | Upper |
|---|---|---|---|---|---|---|---|---|---|
| Step 1[a] | Ages | .723 | .005 | 23858.155 | 1 | .000 | 2.061 | 2.043 | 2.080 |
| | HoldTime | .013 | .000 | 6852.328 | 1 | .000 | 1.013 | 1.013 | 1.013 |
| | FlightTime | .004 | .000 | 3928.928 | 1 | .000 | 1.004 | 1.004 | 1.004 |
| | Constant | -57.434 | .353 | 26467.517 | 1 | .000 | .000 | | |

a. Variable(s) entered on step 1: Ages, HoldTime, FlightTime.

Table 4: Logistic Regression Model

1) Latency time was statistically insignificant.
2) The odds of being affected with PD is 2.06 times higher for every year increase in age (b=.723, p<.01, OR=2.061, (95%CI: 2.043, 2.080)).
3) The odds of being affected with PD is 1.01 times higher for every minute increase in hold time (b=-.013, p<.01, OR=1.013, (95%CI: 1.013, 1.013)).
4) The odds of being affected with PD is 1.004 times higher for every minute increase in flight time (b=.004, p<.01, OR=1.004, (95%CI: 1.004, 1.004)).

### *A. Predict the consequences of PD, based on tremors, hand, and directions:*

Without the independent variables, the model was 57.5% accurate. The prediction power improved to approximately 83.3% after the addition of the independent variables, as shown in Table 5. From the pseudo R2 tests, it was found that the independent variables account for 48% to 64% of the variability of being affected with PD. All these variables were statistically significant with p < 0.05. Note that in this dataset each hand related with specific directions. For instance:

– Right-hand related with Right-Right-Direction, Left-Right-Direction, and Space- Right-Direction.
– Left-hand related with Left-Left-Direction, Right-Left-Direction, and Space-Left- Direction.
– Space key related with Space-Space-Direction, Right-Space-Direction, and Left- Space-Direction.

## VIII. RESULTS INTERPRETATION:

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I.for EXP(B) Lower | Upper |
|---|---|---|---|---|---|---|---|---|---|
| Step 1[a] | Tremors(1) | 5.862 | .044 | 17371.482 | 1 | .000 | 351.335 | 322.007 | 383.335 |
| | Space(1) | 1.569 | .117 | 179.171 | 1 | .000 | 4.802 | 3.816 | 6.042 |
| | Right_Hand(1) | -1.037 | .037 | 783.101 | 1 | .000 | .354 | .330 | .381 |
| | Left_Right_Direction(1) | .654 | .038 | 290.543 | 1 | .000 | 1.924 | 1.785 | 2.074 |
| | Left_Space_Direction(1) | -2.231 | .118 | 354.604 | 1 | .000 | .107 | .085 | .136 |
| | Right_Right_Direction(1) | .694 | .038 | 327.550 | 1 | .000 | 2.001 | 1.856 | 2.157 |
| | Right_Left_Direction(1) | -.382 | .023 | 282.217 | 1 | .000 | .683 | .653 | .714 |
| | Right_Space_Direction (1) | -2.221 | .120 | 344.230 | 1 | .000 | .108 | .086 | .137 |
| | Space_Left_Direction(1) | -.885 | .030 | 891.900 | 1 | .000 | .413 | .389 | .437 |
| | Constant | -.539 | .015 | 1372.478 | 1 | .000 | .584 | | |

a. Variable(s) entered on step 1: Tremors, Space, Right_Hand, Left_Right_Direction, Left_Space_Direction, Right_Right_Direction, Right_Left_Direction, Right_Space_Direction, Space_Left_Direction.

Table 5. Logistic Regression Model

1) Individuals who have tremors as second symptoms are 351.34 times more likely to have Parkinson's than those who do not have tremors.
2) Hand variables: In hand variables my reference group is Left-hand:
   1) When space is typed the odds of Parkinson's is 4.80 times more likely to be affected by Parkinson's than Left-hand (b=1.569, p<.01, OR=4.802, (95%CI: 3.816, 6.042)).
   2) Left-hand in the key presses are 2.82 times more likely to be affected by Parkinson's than Right-hand (b=-1.037, p<.01, OR=.354, (95%CI: .330 , 381)).
3) For Direction variables, I'm going to interpret them based on Left-Left-Direction as a reference group:
   1) A person who does key presses in the Left-Right direction is 1.92 times more likely to be affected by Parkinson's than those who use a Left-Left direction (b=.654, p<.01, OR=1.924, (95%CI: 1.785, 2.074)).
   2) A person who does key presses in the Left-Left direction is 9.3 times more likely to be affected by Parkinson's than those who use a Left-Space direction (b=-2.231, p<.01, OR=.107, (95%CI: .085, .136)).
   3) A person who does key presses in the Right-Right direction is 2.00 times more likely to be affected by

Parkinson's than those who use a Left-Left direction (b=.694, p<.01, OR=2.001, (95%CI: 1.856, 2.157)).

4) A person who does key presses in the Left-Left direction is 1.46 times more likely to be affected by Parkinson's than those who use a Right-Left direction (b=-.382, p<.01, OR=.683, (95%CI: .653, .714)).
5) A person who does key presses in the Left-Left direction is 9.26 times more likely to be affected by Parkinson's than those who use a Right-Space direction (b=-2.221, p<.01, OR=.108, (95%CI: .086, .137)).
6) A person who does key presses in the Left-Left direction is 2.42 times more likely to be affected by Parkinson's than those who use a Space-Left direction (b=-.885, p<.01, OR=.413, (95%CI: .389, .437)).

For example, consider a person who has tremors, presses on the space key in the Right-Space direction. My binary logistic regression model predicts a 61% chance they will have Parkinson's disease.

| Tremors | Space | Right hand | Left Right Direction | Left Space Direction | Right Right Direction | Right Left Direction | Right Space Direction | Space Left Direction |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |

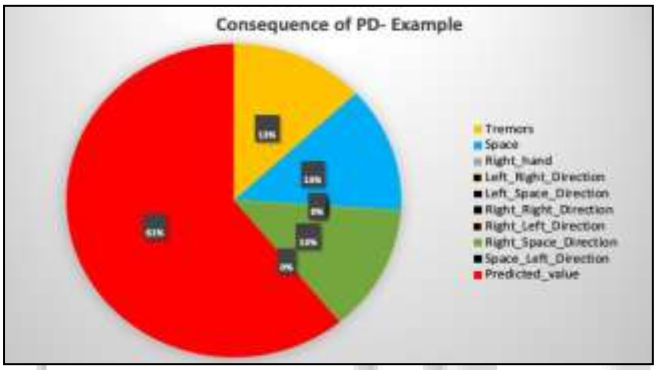


Fig. 3: Third Model - Example

## IX. OUTCOMES:

1) A delay in diagnosing PD increases the probability of the disease progressing.
2) The model used to predict the consequences of PD can also be used to predict the changes in symptoms as the disease progresses over time. This is possible if we know the predictive variables, for example, direction and hand.
3) Model-1 for tremors predicted whether a person would get tremors with approximately 77.8% accuracy.
   – Tremors tend to be less frequent as age increases.
4) Model-2 was able to predict the difference between PD patients and controls with about 94.3% accuracy.
   – Individuals with PD tend to take a longer time when they are typing compared to individuals without PD.
5) Model-3 was able to predict the consequences of PD, based on tremors, hand, and directions, with approximately 83.3% accuracy.
   – As PD progresses, patients will have difficulty when pressing the following keyboard

Keys:

– The space keys.
– Keys on the left side of the keyboard.
– Keys on the right side of the keyboard

## REFERENCES:

[1] C. B. Levine, K. R. Fahrbach, A. D. Siderowf, R. P. Estok, V. M. Ludensky, and S. D. Ross, "Diagnosis and treatment of Parkinson's disease: a systematic review of the literature," Evid. Rep. Technol. Assess. (Summ.), no. 57, pp. 1-4, May 2003.
[2] J. W. Langston, "The parkinson's complex: parkinsonism is just the tip of the iceberg", Annals of neurology, vol. 59, no. 4, pp. 591-596, 2006
[3] G. Rizzo, M. Copetti, S. Arcuti, D. Martino, A. Fontana, G. Logroscino, "Accuracy of clinical diagnosis of parkinson disease a systematic review and meta-analysis", Neurology, vol. 86, no. 6, pp. 566-576, 2016.
[4] D. Heisters, "Parkinson's: symptoms treatments and research", British Journal of Nursing, vol. 20, no. 9, pp. 548-554, 2011
[5] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals (2003). Circulation. 101(23):e215-e220.