### BASIC DETAILS

**1a Title of the proposal**

*Early detection of symptoms of Parkinson’s Disease using finger movements during typing: a machine learning approach.*

### 1b Details of the main applicant

Name: Simon Coopmans

Affiliation:

Address of Knowledge Institute: Department for Human Movement Sciences, University Medical Center Groningen, University of Groningen, A. Deusinglaan 1 9713 AV Groningen

Preference for correspondence in English: **yes**/no

Email: s.c.coopmans@student.rug.nl

**1c Co-applicant(s)**

Name: Martijn van Haren

Affiliation:

Address of Knowledge Institute: Department for Human Movement Sciences, University Medical Center Groningen, University of Groningen, A. Deusinglaan 1 9713 AV Groningen

Preference for correspondence in English: **yes**/no

Email: m.van.haren@student.rug.nl

Name: Thijme Langelaar

Affiliation:

Address of Knowledge Institute: Department for Human Movement Sciences, University Medical Center Groningen, University of Groningen, A. Deusinglaan 1 9713 AV Groningen

Preference for correspondence in English: **yes**/no

Email: T.E.Langelaar@student.rug.nl

Name: Sanne Westerveld

Affiliation:

Address of Knowledge Institute: Department for Human Movement Sciences, University Medical Center Groningen, University of Groningen, A. Deusinglaan 1 9713 AV Groningen

Preference for correspondence in English: **yes**/no

Email: s.westerveld@student.rug.nl

**1d Composition consortium**

|  |  |  |  |
| --- | --- | --- | --- |
| **Consortium** |  |  |  |
| ***Partner(s) research institute(s)*** | ***Institute/organisation*** | ***Role*** | ***Research area*** |
| Dr. M. Kempe | Department for Human Movement Sciences, University of Groningen. | PhD – Scientific support | Data Sceince |
| Dr. W.R. Adams | School of Computing & Mathematics, Charles Sturt University, N.S.W., Australia | Data provider | Life and Health sciences |
| **End user participation** |  |  |  |
| Patients suffering from Parkinson’s Disease |  | Data Provider- End user |  |
| Health practitioners |  | End user |  |

**1e Scientific Abstract**

Parkinson’s disease (PD) is a chronic neurodegenerative disease. The loss of dopamine producing neurons in PD patients results in motor and non-motor symptoms and there is still no cure. Currently, the diagnosis of PD is done by a holistic evaluation of the symptoms of a patient. This means the patient not can be diagnosed before the symptoms are visible and there is already a loss of dopamine-producing neurons before PD has been diagnosed. In addition, misdiagnosis of PD is incredibly high. The goal of this study is therefore to come up with an innovative approach to diagnose PD with more accuracy and at an earlier stage. For this study this research group can use an existing data set. This data set contains anonymous data of health persons and persons diagnosed with PD. From these persons the timing of the key strokes of typing on a computer during their usual computer usage is recorded and stored anonymously and unreadable. A supervised machine learning approach might be possible to build a model to detect early symptoms of PD. A possible way to detect early symptoms of PD might be to differentiate between the kinematics of the keys presses on the left side versus keys pressed on the right side of the keyboard, since onset of the symptoms seem be unilateral. When this research groups achieves this goal, this model can be used in practices to diagnose PD sooner than with the current method. This might lead to PD patients getting medication sooner and thereby reducing and delaying the symptoms of PD.

### 1f Keywords

Parkinson’s Disease, diagnosis, fine motor skills, typing, machine learning

### 1g Scientific Background & Relevance to the research areas:

Parkinson’s disease (PD) is a chronic neurodegenerative disease. The loss of dopamine producing neurons in PD patients results in motor and non-motor symptoms and there is still no cure (Adams, 2017). Currently, the diagnosis of PD is done by a holistic evaluation of the symptoms of a patient (Sveinbjornsdottir, 2016). This means the patient not can be diagnosed before the symptoms are visible and there is already a loss of dopamine-producing neurons before PD has been diagnosed (Fearnley and Lees, 1991). The goal of this study is therefore to come up with an innovative approach to diagnose PD with more accuracy and sooner.

Symptoms in the motor control of PD patients such as bradykinesia, unilateral onset of the symptoms and persistency of these asymmetric symptoms (Sveinbjornsdottir, 2016), might have an effect on a simple typing task. Therefore, it might be possible to use key stroke data of typing on a computer to find differences between healthy and PD patients with mild symptoms.

For this study this research group can use the data of Adams (2017). This data set contains anonymous data of healthy persons and persons diagnosed with PD. This data set includes birth year, gender, presence of several symptoms, medicine usage, and PD severity. From these persons key stroke data on their own computer during their usual computer usage was recorded (using a custom key stroke recording program called ‘Tappy’) without any supervision. From the key strokes the date, timestamp, key on left or right side of keyboard, hold time, latency time, fly time and direction of the key pressing to the next key (left side to right side, left to left side etc.) was recorded.

Misdiagnosis of PD is incredibly high according to Singh and Xu (2019). Methods of data science might help in this case. A supervised machine learning approach might be able to build a model to detect early symptoms of PD. These symptoms might not be visible yet, however differences for instances in key stroke latencies, hold times and flight times might show differences between healthy persons and patients with mild PD.

A possible way to detect early symptoms might be to differentiate between the kinematics of the keys presses on the left side versus keys pressed on the right side of the keyboard, since onset of the symptoms seem be unilateral (Sveinbjornsdottir, 2016). When left hand versus right hand kinematics are not separately analysed, minor differences in for example key stroke latencies, hold times, and flight times between healthy subjects and patients with mild PD might not be detected.

The biggest challenge of the study is to find a model to that can differentiate between healthy and PD patients with mild symptoms with high accuracy and precision. When this research groups achieves this goal, this model can be used in practices to diagnose PD sooner than with the current method. This might lead to PD patients getting medication sooner and thereby reducing and delaying the symptoms of PD.

Data Science:

Sport Science:

Life Sciences and Health:

Creative Industry Research:

Social sciences and/or humanities (ethics and legislation):

### 1h Research Questions & Approach

### 1. What aspects of typing data can be used to make a distinction between healthy persons and Parkinson patients?

### 2. Can Tappy data be used to detect Parkinson disease in an early stage?

### This proposal is a continuation of earlier research and data collection of keyboard data in elderly persons. The data is available from 103 participants, this will be sufficient to start machine learning on the data. The data is collected with Tappy and is clearly structured. There are 3 options possible on which a distinction can be made. HoldTime x Hand gives 3 features, Time x Directions gives 27 features. Besides that time of the day can also be used to make a distinction. In the start of the project we will feed the machine learning bits of the data that is clearly specified as either being Parkinson patient or a healthy person. A unique distinction will be found out of those previous mentioned features and then unlabeled data will be given to the program. Ideally in the end a 100% score will be reached where every Parkinson patient is separated from the healthy adults.

### Data analysis procedure

### All the data will be tested on multiple features and on multiple interactions to have a broad spectrum to get the best distinction between healthy and mild Parkinson patients. Further, the effect of medication on the result of the model also needs to be taken into account. The exact order of all the features that will be tested cannot be foreseen. At the end the best detection method will be chosen and will be used to make an app that can be used by people and doctors to detect Parkinson disease.

Word count: [324]

### 1i Specification of the Big Data and/or Data Science aspects

Collecting typing data gives a lot of data over a short period of time. Also a lot of information of 1 key stroke is saved. This means that the volume and variety will be very big. As is said a bit below the timing is also very important, thus the velocity of the data needs to be high in order to get the right timing. Big veracity is also a problem that can occur with the typing data. How do you make a distinction between someone typing slow deliberately and a Parkinson patient.

The application ‘Tappy’ for Windows computer collects all the data, for every keystroke it collects the following data:

* Time of the day: Which gives the time of the day at which each keystroke begins accurate to a 1/100st of a second.
* Hold time: The time between key down and key up in milliseconds. It gives the speed of the tapped finger.
* Latency: Time in milliseconds between key down of one key to key down of the next key. Distinction is made between subsequent keystrokes with the same hand or with the other hand. A latency bigger then 800ms is considered a pause in typing.
* Flight time: Time between key up and the next key down.

The dataset contains only data from participants with more than 2000 keystrokes. All this data is completely anonymised. Given all the data that is collected per keystroke and the amount of keystrokes that each participant has the data needs a good structure. A Big Data approach will give the best structure to the data and also to implement the data the best for the future.  
With machine learning the data will be analyzed to search for a distinction between mild Parkinson Disease and healthy participants. Big Data is good to use for this because in the future the amount of data will only be more and bigger. If the machine learning model is already set up for Big Data the bigger data sets in the future won’t be a problem.

Word count: [338]

**1j End-user participation**

### In this research all the data is already collected but in the data collection the participants had a visit from the researcher. The researchers downloaded Tappy for them and the data was collected on the computer of the participant self. Every now and then the data was send to the researchers. Because on forehand it was known if a participant had Parkinson disease or not there was not any feedback send to the participants. The research was just done to train the Machine learning program.

Word count: [85]

**1k Innovations**

### If via machine learning the detection of Parkinson disease can be made a lot of possibilities will come available. Detection can be done in an earlier stage of Parkinson disease. Less patients will be wrongly diagnosed. Doctors will have a good test method to detect Parkinson disease. An app for the computer can be created that can be downloaded by people themselves. This way people can check themselves if they have Parkinson and then go to the doctor. The app could also be expanded to check on the progress of Parkinson disease in people who are already diagnosed. Detection of other disease can be done via typing data. In the future Parkinson disease can be detected at home. Also the timing of intake of leva dopa could be done using this method. When the keyboard detects that the person has a harder time typing it could give a signal that it is time to take leva dopa. This way the dose could be more administered at the moments the patient really needs it.

### 

Word count: [165]

**1l Work description**

|  |  |  |
| --- | --- | --- |
| ***Work packages*** | ***Personal Involved***  ***(Name, Extern/ Intern)*** | ***Expected Hours*** |
| 1 Data Collection | W.R. Adams - Extern | 50 |
| 2 Data preprocessing | W.R. Adams - Extern | 150 |
| 3 Buidling Pattern Recognition Algorythm | Master students - Intern | 100 |
| 4. exploratory analysis | Master students – Intern | 50 |
| 5. Classifier 1 with all data | Master students - Intern | 50 |
| 6. Classifier 2 with all data | Master students – Intern | 50 |
| 7 tradidtional Analysis | Master students- Intern | 50 |
| 8 Machine learning Analysis | Master students- Intern | 200 |
| **Total** |  | **750** |
| **Ratio\* intern/ extern research** | [X] % versus [Y] % (total = 100%) | **73 – 27 %** |

*\* defined by the amount of requested funding per work package.*

### FINANCIAL BUDGET

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***Project budget*** | ***Cash*** | ***In Kind*** | ***Description*** | ***Total*** |
| Senior Researcher | 15000 |  | The Senior Researcher helps with the whole project and guides the master students. | € 15000 |
| Master Student |  | 2500 | The master student must do their graduation project on this study, which is 40 ECTS | € 2500 |
| Master Student |  | 2500 | The master student must do their graduation project on this study, which is 40 ECTS | € 2500 |
| Master Student |  | 2500 | The master student must do their graduation project on this study, which is 40 ECTS | € 2500 |
| Master Student |  | 2500 | The master student must do their graduation project on this study, which is 40 ECTS | € 2500 |
| Travel expenses | 3300 |  | 4 tickets for the master students to Nice which are 100-140 euro per person. Hotel is booked via congress organization. The congress itself is the International congress of Parkinson’s Disease and movement disorders. 1 of the three topics is: ‘Discuss the diagnostic approaches and tools available for Parkinson’s disease’. | € 3300 |
| Poster print | 30 |  |  | € 30 |
| Data acquisition |  | 5000 |  | €5000 |
| Data Base |  | 5000 |  | 5000 |
| Data Base Design |  | 8000 |  | 8000 |
| ***Total*** | ***48330*** | ***28000*** |  | ***€ 76.030*** |
|  |  |  |  |  |
| **Percentage Matching** |  |  |  |  |

Travel expenses is for the International congress of Parkinson’s Disease and movement disorders in Nice, France where 1 of the 3 topics is: ‘Discuss the diagnostic approaches and tools available for Parkinson’s disease’.

### References

Adams, W. R. (2017). High-accuracy detection of early parkinson's disease using multiple characteristics of finger movement while typing.*PloS One, 12*(11), e0188226.

Fearnley, J. M., & Lees, A. J. (1991). Ageing and parkinson's disease: Substantia nigra regional selectivity.*Brain : A Journal of Neurology, 114 ( Pt 5)*(5), 2283-2301.

Singh, S., & Xu, W. (2019). Robust detection of parkinson's disease using harvested smartphone voice data: A telemedicine approach.*Telemedicine Journal and E-Health : The Official Journal of the American Telemedicine Association.*

Sveinbjornsdottir, S. (2016). The clinical symptoms of parkinson's disease.*Journal of Neurochemistry, 139 Suppl 1*, 318-324.