### 1h Research Questions & Approach

### *## Name the main research question of your proposal and the basic approach to answer these questions. Divide the description of your approach into an general description and the data analysis process (procedure)*

### *## Use up to 400 words.*

### 1. Detecting Parkinson disease in an early stage

### 2. Can keyboard 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. The participants were completely anonymous and they just did their regular typing, no typing task was provided. Only the letters were recorded, not the numbers. The only key that can be really seen in the data as a separate key is the space bar. The other keys are ranked based on the column they are in and the hand that is used to type them. So the keys W,S,X are ranked as column 2 left hand, a distinction between these 3 keys cannot be made. This is all to keep the data anonymous and unreadable. Known already is that pure hold time, latency time and flight time will not be enough to make a distinction between healthy elderly participants and persons with mild Parkinson disease. This study will look into the possibilities to detect Parkinson disease based on interactions between Hold time and Hand, Times and Direction. Also Time of the day will be used to get a better detection of Parkinson disease.

### 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.

Data analysis procedure

For the first and second research question will we use a stepped approach in data analyses. This stepped approach is designed to start with existing knowledge approaches (Step 1) but integrating richer data sets. In the second step, we will apply, state of the art data science, approaches that are not yet commonly adopted in human movement science. The project will be innovative for human movement science, exercise science literature, and for applications of data science. The requested data science support will therefore be asked during step 2 only.

Step 1: Analyses will be performed with traditional approaches (e.g. regression and mixed model´s).

Step 2: Supervised, Recurrent Neuronal Networks and convolutional neuronal networks approaches will be explored to detect movement pattern in the combined data set (GPS and Accelerometer). Furthermore, supervised approaches (Support-Vector-Machines, Recurrent Neuronal Networks) and unsupervised approaches (Self-Organizing Maps) will be explored to capture the association between movement pattern and internal load variables.

Step 3: Evaluation of the different approaches and integration of the most appropriate combination to create a precise model to determine the internal load.

Word count: [xxx]

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

### *## Motivate why this project concerns Big Data*

### *## Describe the character of the (future) data sets to be used*

### *## Describe the added value or necessity of the (future) usage of Big Data in this project proposal*

### *## Use up to 300 words.*

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. Big Data is needed to collect and process all the data.

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: [281]

**1j End-user participation**

*## Describe how the participation of end-users (patients, citizens and/ or (health) professionals) is organized throughout the project.*

### *## Use up to 250 words.*

### 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**

*## Describe the innovative character of our research on possible innovative outcomes*

### *## Use up to 250 words.*

### 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. People who are in doubt themselves could easily download the app in the future to test if they have Parkinson disease. The app could also be expanded to check on the progress of Parkinson disease in people who are already diagnosed. Maybe detection of other disease can be done via typing data. In the future Parkinson disease can be detected at home. Maybe 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**

*## describe all the work packages, hand a Gantt Chart as appendix at the end of the proposal*

|  |  |  |
| --- | --- | --- |
| ***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 Designing Data Base | ???? - ???? | 50 |
| 4 Buidling Pattern Recognition Algorythm | WE - Intern | 100 |
| 5 tradidtional Analysis | WE- Intern | 50 |
| 6 Machine learning Analysis | WE- Intern | 150 |
| **Total** |  | **??** |
| **Ratio\* intern/ extern research** | [X] % versus [Y] % (total = 100%) | **58 – 42 %** |

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

### FINANCIAL BUDGET

*##* *Specify for the project the requested budget for personnel and project-/specific costs,*

*##* *the contribution by private and/or public partners and specification in in-kind and in-cash.*

*## The budget should express a justifiable balance between targets and available resources (budget, staff, traveling, and resources).*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***Project budget*** | ***Cash*** | ***In Kind*** | ***Description*** | ***Total*** |
| Senior Researcher | 15000 |  |  | € 15000 |
| PhD Student | 30000 |  |  | € 30000 |
| Master Student |  | 2500 |  | € 2500 |
| Travel expenses | 1500 |  |  | € 1500 |
| Poster print | 40 |  |  | € 40 |
| Data acquisition |  | 1000 |  | €1000 |
| Data Base |  | 5000 |  |  |
| Data Base Design |  | 8000 |  |  |
| ***Total*** | ***46540*** | ***16500*** |  | ***€ 63.040*** |
|  |  |  |  |  |
| **Percentage Matching** | 73.8 | 26.2 |  |  |

### References