|  |  |  |  |
| --- | --- | --- | --- |
|  | Name: John Fitzgerald  StudentID: R00156081  Email : john.p.fitzgerald@mycit.ie |  |  |
| Journal 1 | | | | |
| Project Summary | | | | |
|  | | | | |

|  |  |  |
| --- | --- | --- |
| Report Date | Project Name | Prepared By |
| 12-02-2022 | Classification of Major Depressive Disorder (MDD) Using Motor Activity Data | John Fitzgerald |

|  |
| --- |
| Description and main objective |
|  |

Studies show that mobility in people with behavourial disorders is lower than that of healthy individuals [1]. It is reported that 80% of schizophrenic patients can suffer activity disruption while they are asleep, and also during the day [2,3]. There is also a high percentage of insomnia in adult patients suffering from depression.

This project will use activity data, recorded on wrist wearable devices, gathered from Depressive, Schizophrenic and Healthy subjects over a period of time. I will clean, process and analyse the data before it will be classified using appropriate algorithms for the type and volume of data. Previous work on this topic will be examined and it is hoped that this project can improve and enhance on those earlier project’s results.

|  |
| --- |
| DATASET |
|  |

Two independent datasets which have been published by the same research group will be used [ https://datasets.simula.no/ ]. The ‘*Psychose*’ dataset which comprises of 22 subjects diagnosed with Schizophrenia and the ‘*Depresjon*’ dataset which consists of 23 subjects suffering from a depressive disorder. There is also a dataset of 32 healthy individuals which will be used as control data.

I propose to merge the 3 datasets to create a file containing readings of activity recorded for 77 individuals over a common period of days. Initial inspection of the data shows the following:

* 23 .csv files, each containing an individual’s activity readings per minute (diagnosed depressive)
* 22. .csv files, each containing an individual’s activity readings per minute (diagnosed schizophrenic)
* 32 .csv files, each containing an indiviidual’s activity readings per minute (diagnosed healthy)

A manual inspection of the files shows a varying numbers of minute readings (number of records) in each CSV – which indicates that the trials may not have adhered to strictly measuring each minute of the day over the 13 day cycle (13 cycle has been used in previous projects using this data). There are 10080 minutes in 13 days. All manually examined files seem to have recorded generally, more than 13 days of recordings. Further analysis will be performed to use the optimal number of days with 24 hours coverage common to all 3 categories.

|  |
| --- |
| RESEARCH GAP AND SOURCES |
|  |

Behavioural and depressive disorders are hard to diagnose. Clinical diagnosis for both disorders are predominantly subjective and is based on self-reporting or observation by a trained physician. Machine Learning presents an ideal opportunity to investigate if the disorders can be identified and classifed based on recorded data of diganosed sufferers and a control group of healthy individuals.

Depression is measured by the Montgomery-Asberg Depression Rating Scale (MADRS). Based on observation and discussion with the patient, 10 relevant items for depression are scored. 0-10 scores are classified as an absence of depressive symtoms and scores above 30 indicate severe depression.

There is a opportunity to build on previous projects that have used Machine Learning for depression states classification. There is an opportunity to detect depression states based on the data provided and also to predict MADRS scores based on motor activity data. Sleep pattern analysis on depression/schizophrenic/control participants will be analysed also and classification will be attempted on sleep patterns.

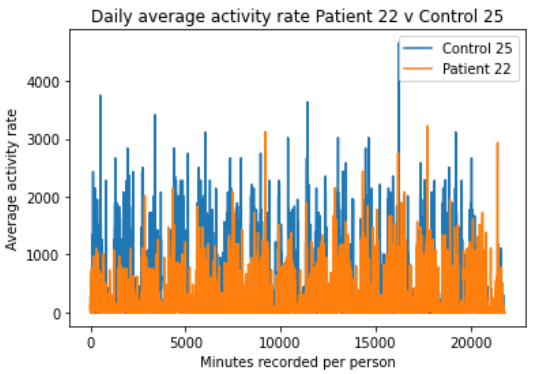
|  |
| --- |
| METHODOLOGY |

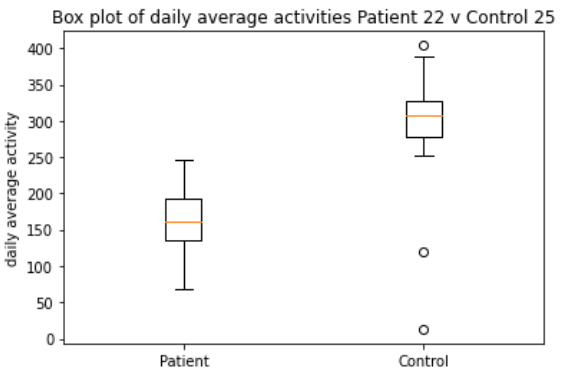
**Schizophrenia dataset:**

The Psychose zip contains python code (*psykose\_baseline\_experiment.py*) which performs experiments on the Shizophrenic data using four different algorithms – Logistic Regression (LR), Random Forest (RF), Extreme Gradient Boosting (XGB) and Light Gradient Boosting (LGB). Amendments to the code were required as versions of packages, datasets or code may have changed. The 4 algorithms report the average precision (PRC) and the area under curve (ROC). All of the algorithms perfomed well in testing. Average PRC and ROC for all four was greater than .80. My results differed slightly to the those in paper ‘PSYKOSE: A motor activity database of patients with schizophrenia’. This may have been to changed in the datasets or code – this needs further investigation. Table II in that paper contains data which is wrongly labelled: it describes PRC results for all four algorithns and Ensemble as ROC and visa-versa.

**Depression Dataset**:

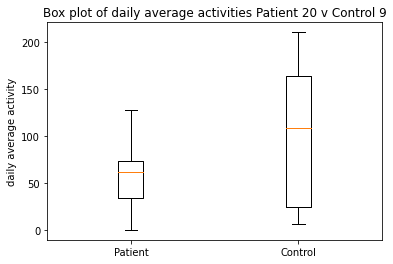
The depression package contains information on all participants (*scores.csv*). It contains useful information on both Patients and Controls. This allowed a preliminary comparision of individual control and patient average daily readings with the same gender and age grouping to compare activity. Figure 1 shows the comparison of Patient 22 and Control 25. Both are males, aged 65-69. Patient 22 has a reading of 29 on the MADRS scale (scores above 30 on this scale indicate severe depression). It is clear from the results that Control 25 has a much higher level of activity.This is demonstrated also in Figure 2 which shows a boxplot of activity for both indviduals, while the spread is much wider for control 25 – there are outliers, but still an obvious higher median of activity compared to Patient 22.

**Figures1&2:**



# 

**Figure3&4:**

Figure 3 (above) is a comparison of Males aged 30-34 where patient 20 was diagnosed as 27 on the MADRS scale. Control 9 shows a higher level of activity for the duration of the test. Figure 4 demonstrates again a wider spread of activity with a higher median for Control.

|  |
| --- |
| MONTHLY PROGRESS EXPECTED |

It is proposed that the datasets will be merged into the following structure (figure 5) to allow data to be concatenated, standardised, transposed and grouped by different time periods. This is subject to change pending further data analysis.

***Fig.5 Proposed cleaned and merged file structure:***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **PERSONID** | **GROUP** | **DAY** | **DATE** | **TIME** | **ACTIVITYREADING** |
| **1-77** | **Depressive/Schizphrenic/Control** | **1-n** | **xx/xx/xx** | **xx:xx** | **0-10000** |

Using Python 3.11.1 to merge, clean and edit the source actigraphy data, Summary statistics will be generated from the data in daily, daily sections (Morning,afternoon,evening,night) and hourly segments. This is to distignuish patterns between the 3 categories of participants.

I will be performing regression analysis on the the data to examine correlation between the groupings and also within 24 hour segments.

There is an opportunity to test other algorithms to optimize prediction results:

* Support Vector Machines: Used a lot in sleep-wake analysis
* Artifical Neural Networks: Deep learning algorithms that train a neural network to predict a categorical outcome
* Convoluted Neural Networks: Deep learning models that are very good at analysing signals
* Long short-term Memory Networks: designed also to handle sequential data which makes it well suited to time series analysis

References:

1. Cosgrave, J.; Wulff, K.;Gehrman,P. – Sleep, circadian rhythms, and schizophrenia: Where we are now and where we need to go. Curr. Opin. Psychiatry **2018**,31, 176-182.
2. Vahia, I.V.; Sewell, D.D. Late-life depression: A role for accelerometer technology in diagnosis and management. Am. J. Psychiatry **2016**, 173, 763–768.
3. Hombali, A.; Seow, E.; Yuan, Q.; Chang, S.H.S.; Satghare, P.; Kumar, S.; Verma, S.K.; Mok, Y.M.; Chong, S.A.; Subramaniam, M.

Prevalence and correlates of sleep disorder symptoms in psychiatric disorders. Psychiatry Res. **2019**, 279, 116–122.