Diagnosing Alzheimer's through Handwritten Tasks

By Elizabeth Bekele

Introduction

Alzheimer’s has been around since the early 1900s.

As Alzheimer’s progresses it debilitates a person’s ability to do daily tasks, increases memory loss, and deteriorates their nerve cells.

A cure for disease has yet to be found. Although there is no cure, there is treatment that can reduce the amount of brain damage and improve the quality of life of those affected by it. This is why it is imperative to detect it early.

Problem

Alzheimer's is Increasing

The human lifespan is lengthening, therefore it is expected that the number of cases will increase as well.

No Cure

Since there is no cure there is a need for improved methods to detect the disease earlier.

Few Information

Previous studies have failed to gather a large dataset with agreeable features to represent their findings making their study weak.

Approach

Inspect a subset of tasks (5 out of 25)

Task 3 (G): join two points with a vertical line, continuously for 4 times

Task 8 (C): write cursively a sequence of four lowercase letter ‘l’ in a single smooth movement

Task 13 (C): copy the word “mamma” above a line

Task 18 (M): write the name of the object shown in a picture (a chair)

Task 23 (M): write a telephone number under dictation

Inspect a subset of features (15 out of 18)

Time to Complete Tasks: tasks on air, paper, and total tasks

Paper Tasks: Mean speed, acceleration, and jerk on paper

Air Tasks: Mean speed, acceleration, and jerk in air

Tremor: Generalization of Mean Relative Tremor (GMRT) on paper, in air, and average GMRT

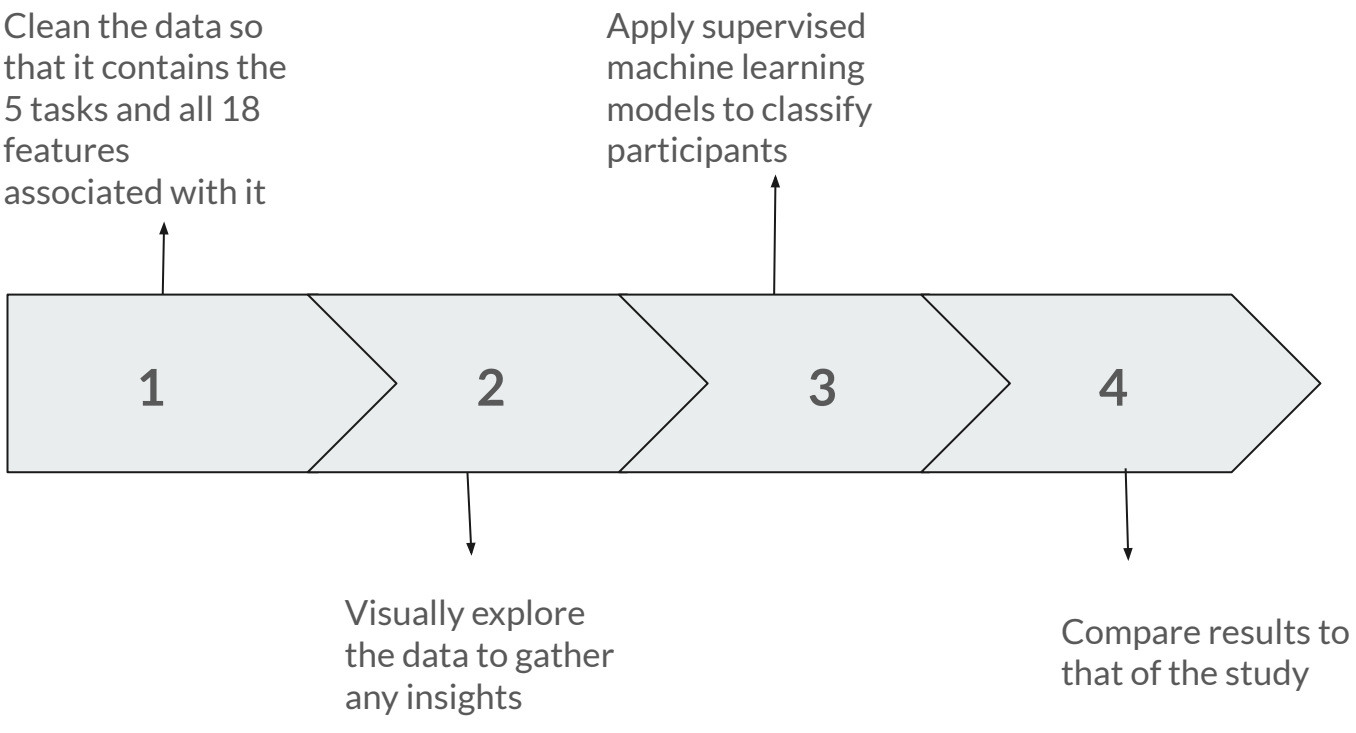
# of Pendowns: counts the # of continuous uninterrupted lines while completing a task

Pressure: variance of pressure levels exerted by the pen tip

Are the tasks & features effective?

- To determine this we will use 4 different machine learning techniques to see if these tasks are effective in classifying participants as either with Alzheimer's or not
- Further, we will compare the results of our models to that of those listed in the study

Implementation

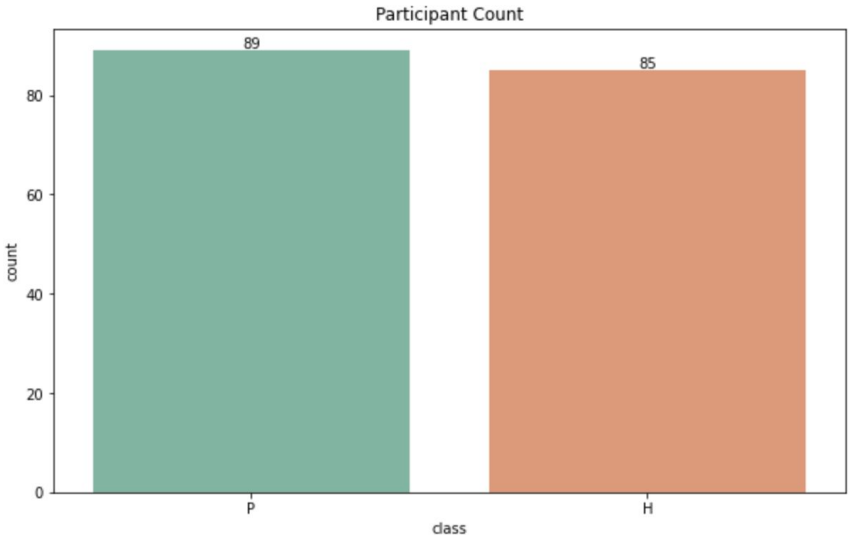


Exploratory Data Analysis

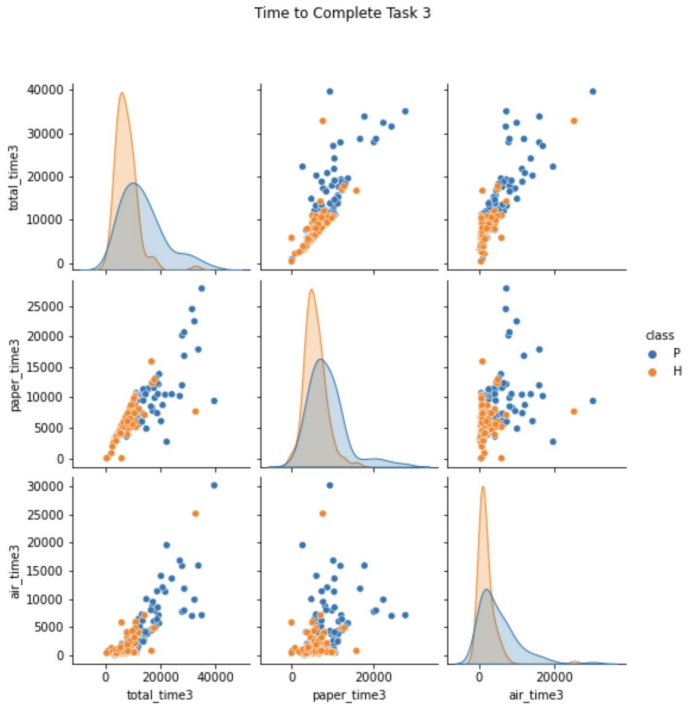
Participant Count

Of the 174 participants, 89 have Alzheimer's and 85 are healthy individuals.

The data is relatively balanced which will allow us to use accuracy as a metric to evaluate our machine learning models later on



Time to Complete a Task



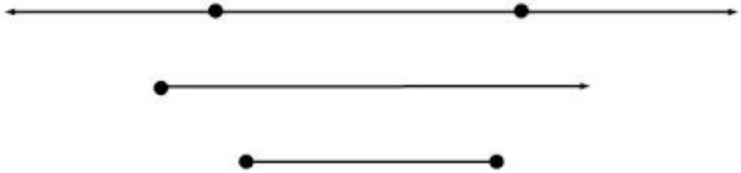
Compared to healthy participants, those with Alzheimer's took longer to complete tasks despite if they were written on paper or conducted in the air.

The same is true for the remaining 4 tasks.

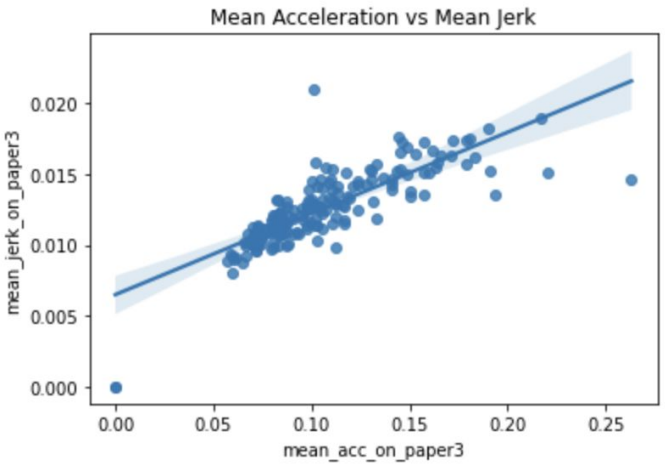
Task 3

Join two points with a vertical line, continuously four times

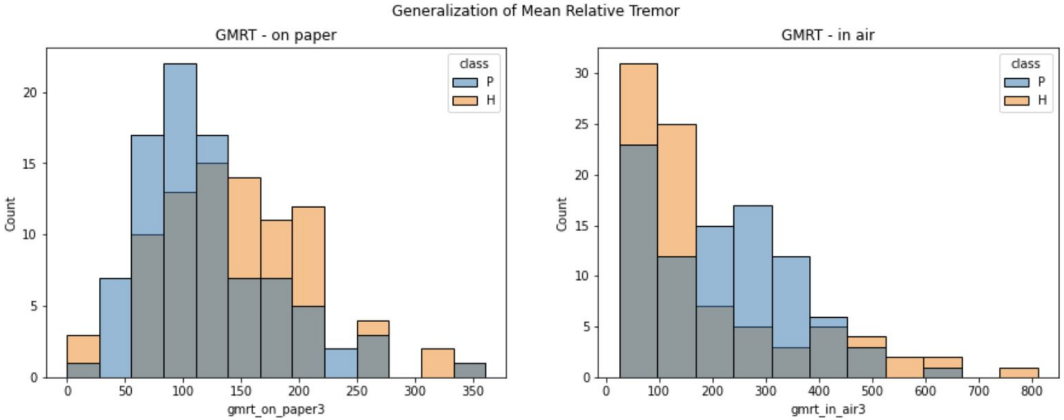
Classified as a **Graphic Task,** used to test participant's ability in writing elementary traits



Task 3 Takeaways



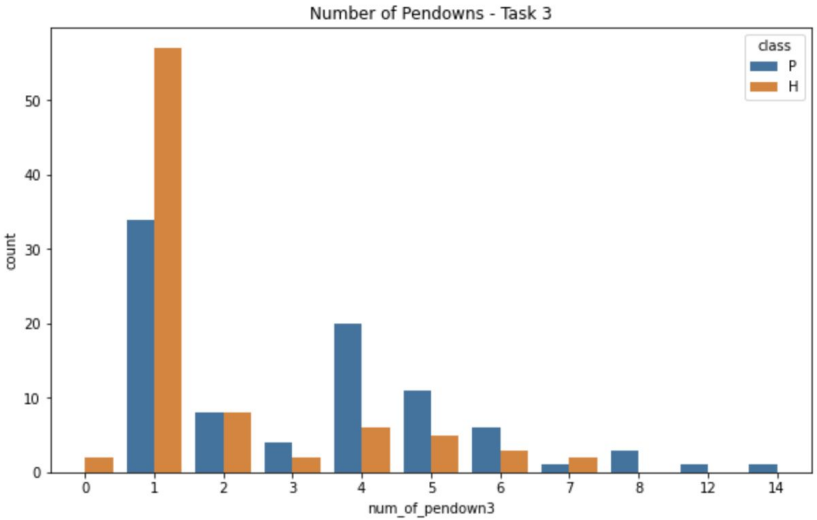
Strong positive relationship between the average acceleration and jerk on paper. This is expected because jerk calculates the rate of change in acceleration



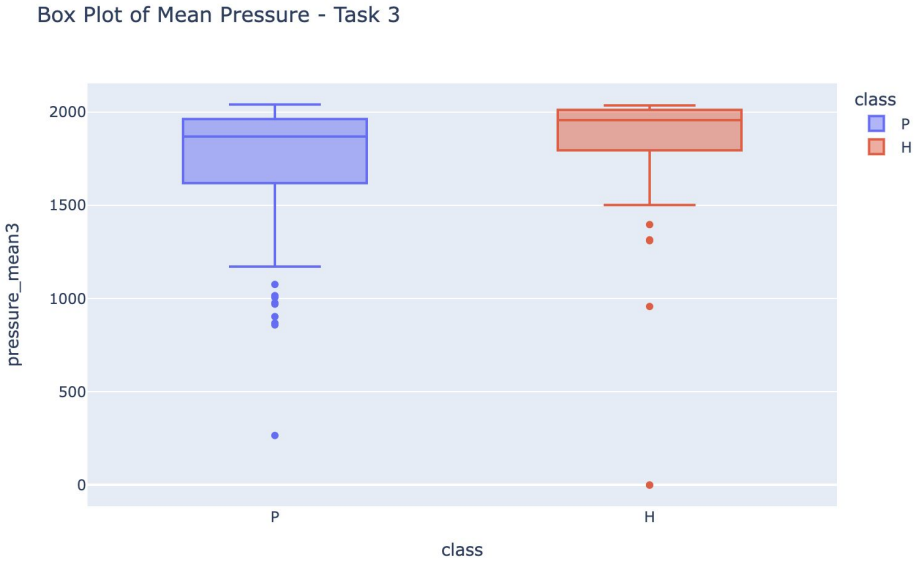
Higher frequency of tremors in the air as opposed to on paper

Tremors in air distribution is heavily right skewed.

Task 3 (cont.)



Participants with Alzheimer's had a higher frequency in the number of pendowns. It's possible that the participant would lose their train of thought leading to the higher frequency.



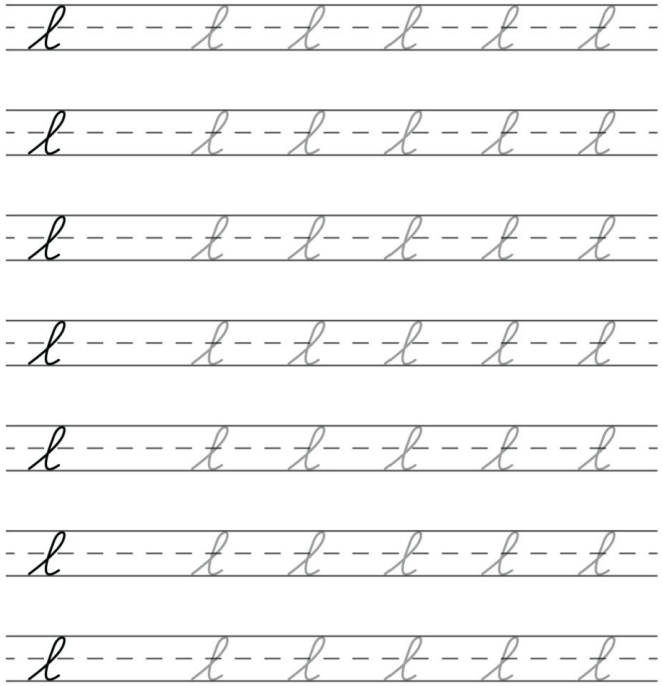
Alzheimer's: 75% of the average pressure exerted falls below 1962.122

Healthy: 75% of the average pressure exerted falls below 2,011.67

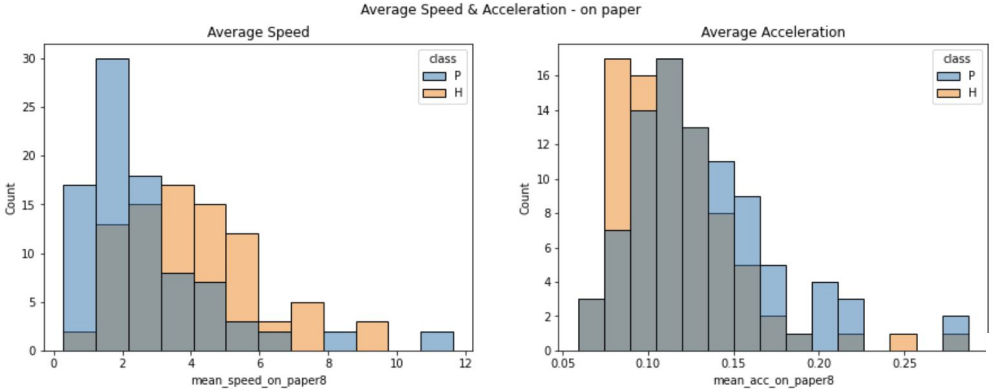
Task 8

Write cursively a sequence of four lowercase letter ‘l’ in a single smooth movement

Classified as a **Copy Task,** evaluates participant’s abilities in repeating complex graphic gestures



Task 8 Takeaways

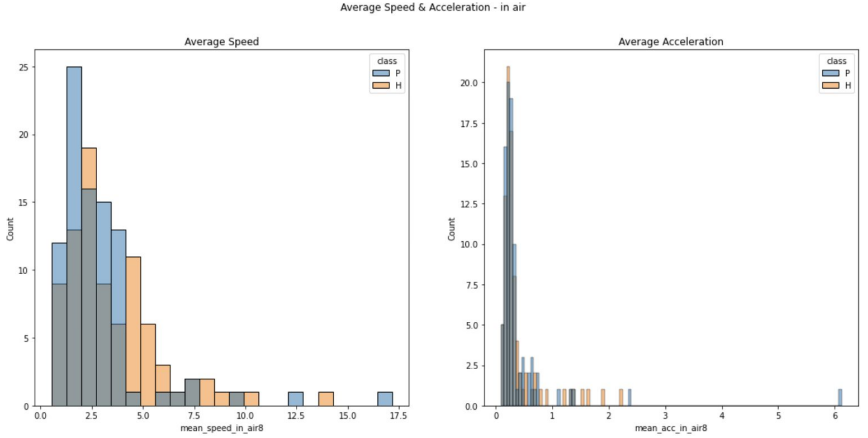


The spread for acceleration on paper is narrower compared to the spread of acceleration in the air.

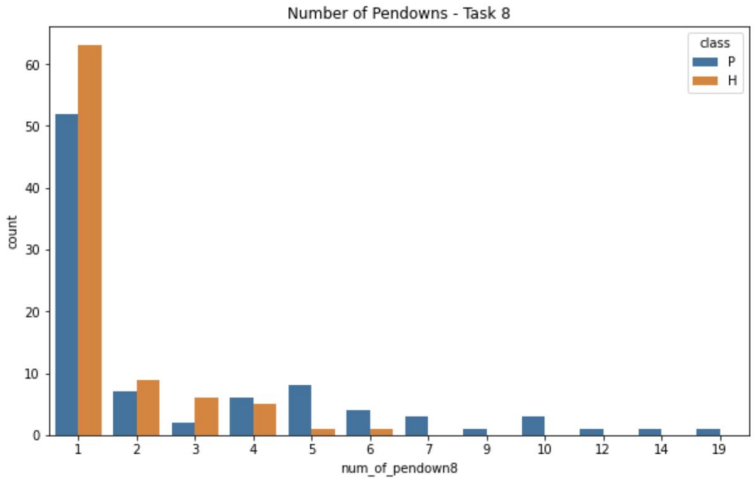
0.05-0.3 (on paper) vs 0-6 (in air)

The speed on paper for healthy patients is centered from 2-6. The speed on paper for patients is centered from 0-4.

The speed in air for healthy patients is centered from 0-6. The speed in air for patients is centered between 0 and 3.

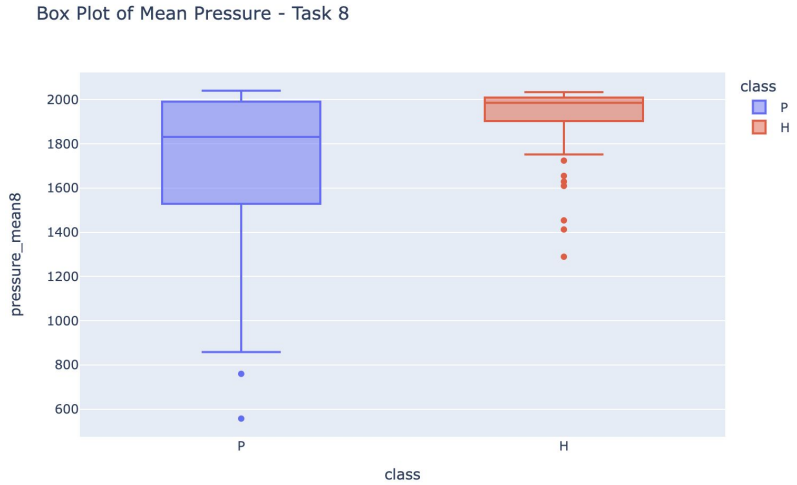


Task 8 (cont.)



Patients with Alzheimer's have a wider spread in the number of pendowns as opposed to healthier individuals.

The range of number of pendowns for healthy participants go from 1-6; however, 1-19 for those with Alzheimer's



Alzheimer's: 75% of the average pressure exerted falls below 1990.77

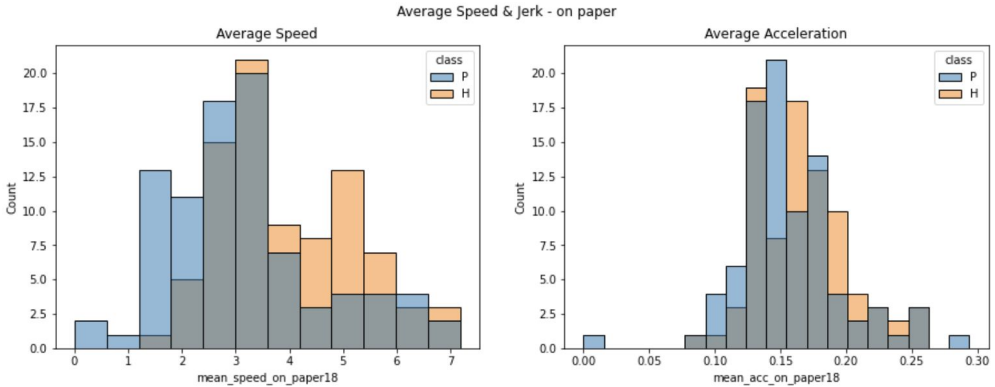
Healthy: 75% of the average pressure exerted falls below 2009.658

Task 18

Write the name of the object shown in the picture (a chair)

Classified as a **Memory Task**, tests the changes in writing process previously memorized or associated with objects shown in a picture

Task 18 Takeaways



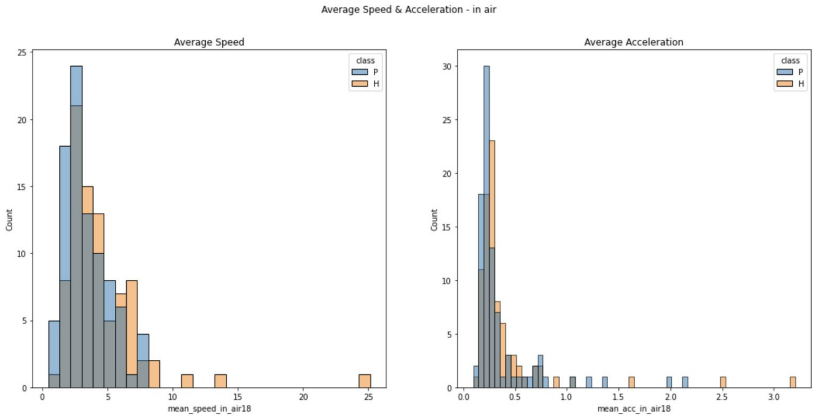
On paper:
The average speed is relatively the same among those with Alzheimer's and those who are healthy.

The average acceleration is concentrated between 0.10 and 0.25 for both classes with the exception of 2 participants with Alzheimer's

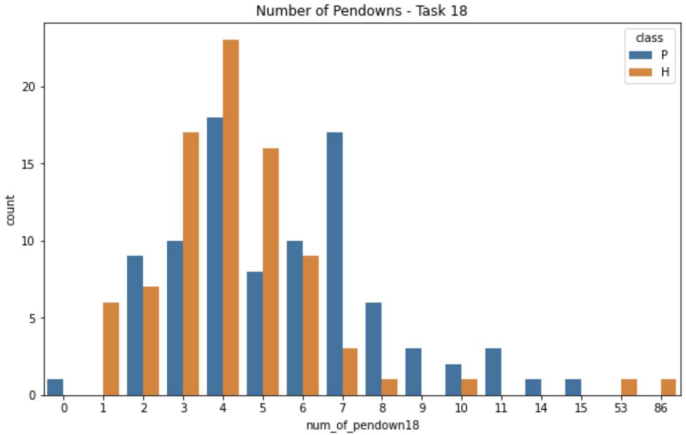
In air:
Both measurements (speed & acceleration) are right skewed.

The average speed is concentrated between 0 and 7 for both classes with the exception of 3 healthy participants.

The average acceleration is concentrated between 0.1 and 0.8. There are outliers in both classes for the average acceleration.

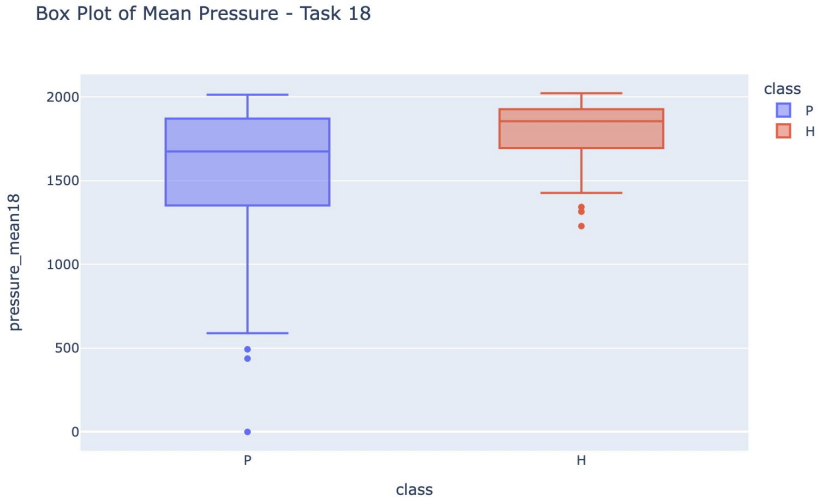


Task 18 (cont.)



There are 2 outliers for the Healthy class

When omitting these outliers, the range for number of pendowns is 0-15 for those with Alzheimer's and 1-10 for those considered healthy



Alzheimer's: 75% of the average pressure exerted falls below 1870.632

Healthy: 75% of the average pressure exerted falls below 1972.035

Machine Learning Models

Logistic Regression & Random Forest

Logistic Regression

- Used for classification
- Estimates the probability that an instance belongs to a certain class

Random Forest

- Used for classification and regression. Here it will be used for classification
- Composed of a group of decision trees.
- Randomizes both the training data instances and features
- Predictions in classification problems are determined based on majority vote

Support Vector Machine & Artificial Neural Networks

Support Vector Machine

- Used in both classification and regression. It will be used for classification here
- Applicable to boundaries that are linear or nonlinear
- *Linear boundaries:* looks for the widest margin between the two classes. Requires all data instances to be on the correct side of the margin (hard margin)
- *Nonlinear boundaries:* allows some margin violation (soft margin)

Artificial Neural Networks

- Composed of several layers of neurons
- Feed-forward network
  - Data moves from the input layer → hidden layers → output layer
- Pre-activation and activation functions are applied to the neurons in the hidden layer
  - Pre-activation consists of the weighted sum of inputs and bias
  - Activation function used here will be sigmoid because this is a binary classification problem

Model Evaluations on all 5 Tasks

Model Evaluations – All 5 Tasks

	Model	Accuracy
0	Logistic Regression	0.79
1	Random Forest	0.89
2	Support Vector Machine	0.74
3	Neural Network	0.79

Random Forest performed the best in classifying participants as either Healthy or with Alzheimer's.

Model Evaluations from Study – All 5 Tasks

	Model	Accuracy Ranges
0	Logistic Regression	74.66 - 89.06
1	Random Forest	83.39 - 93.19
2	Support Vector Machine	71.45 - 86.55
3	Neural Network	75.17 - 91.11

All 4 machine learning algorithms performed within range of those listed in the scientific study.

Thank you!