Walking Analysis with Machine Learning

GA DSIF 9 Capstone Project

Prepared by: Zheng Ting

Prepared on: 23 May 2023
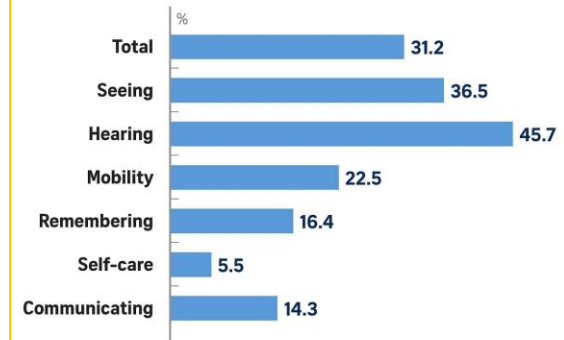
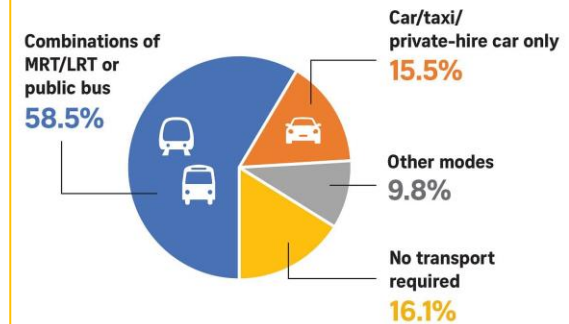# Background and Problem Statement

Problem Statement

1. In addition to Singapore's ageing population, there are many factors affecting mobility and quality of life, such as loss of muscle with age, osteoarthritis and injuries.

2. Among 150,000 household surveyed in 2020 population census, 62,500 people faced mobility challenges such as walking and climbing steps.

Labour force participation among residents aged 15 to 64 with difficulty performing basic activities

%

Total	31.2
Seeing	36.5
Hearing	45.7
Mobility	22.5
Remembering	16.4
Self-care	5.5
Communicating	14.3

Mode of transport to work among residents aged 15 and over with difficulty performing basic activities

Combinations of MRT/LRT or public bus **58.5%**

Car/taxi/ private-hire car only **15.5%**

Other modes **9.8%**

No transport required **16.1%**

NOTE: Data excludes working persons who were overseas for at least 6 months.

Source: DEPARTMENT OF STATISTICS
STRAITS TIMES GRAPHICS

Objectives

1. Using machine learning to identify key features that relates to different walking issues between Healthy Controls, Hip, Knee and Ankle respectively.

2. Explore possibility of affordable wearable sensors that target these key features for general public to monitor outcomes of treatments.
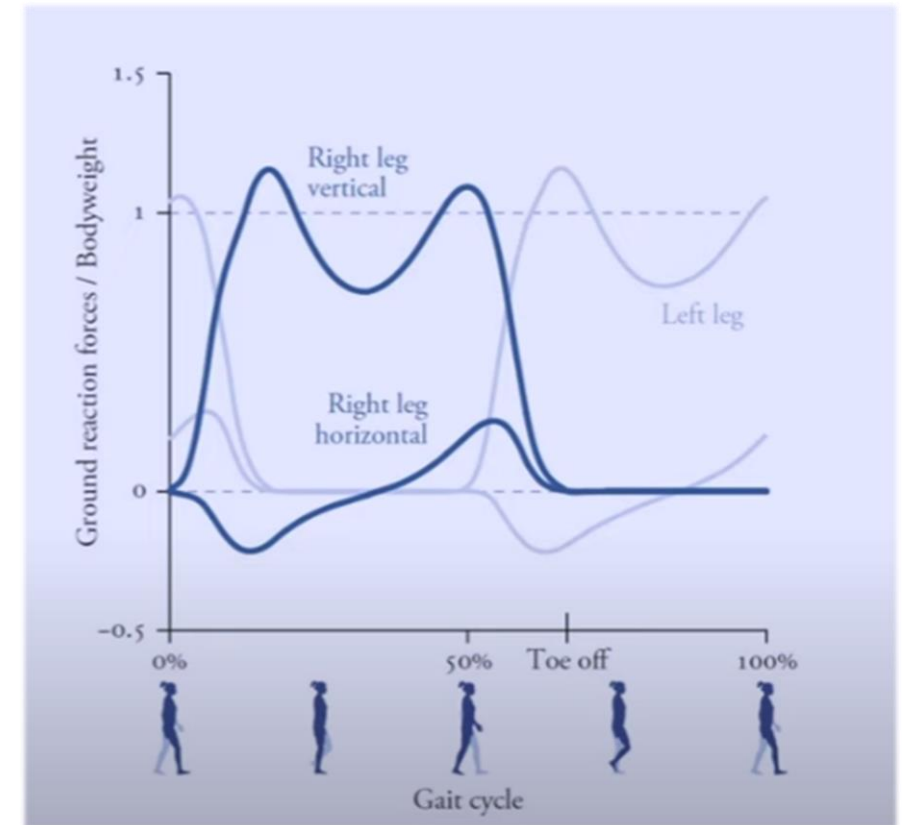

Source: https://americanbonehealth.org/exercise/the-mechanics-of-walking/
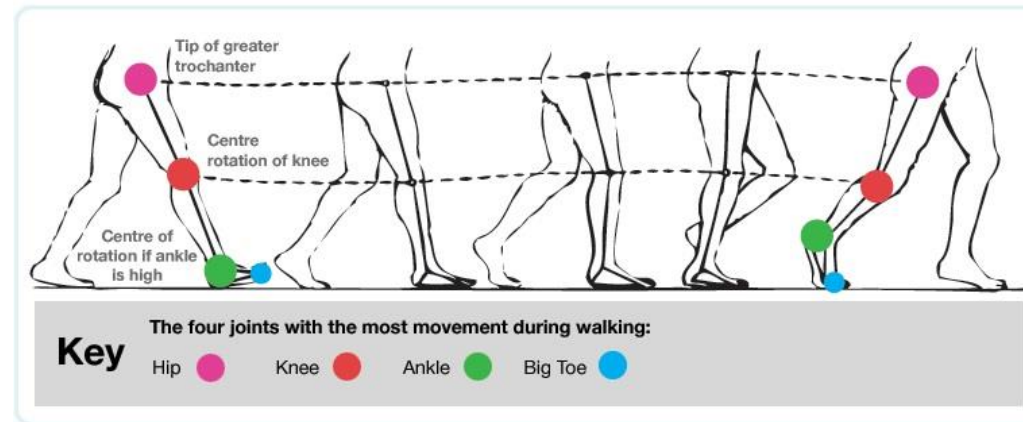
Measurements

There are 3 types **Ground reaction forces** (N):

- Vertical (Landing and push-off)

- Horizonal (Braking and propulsion)

- Rotation

There are 2 types of **Center of Pressures** (Cm):

- Horizontal (Braking and propulsion)

- Rotation

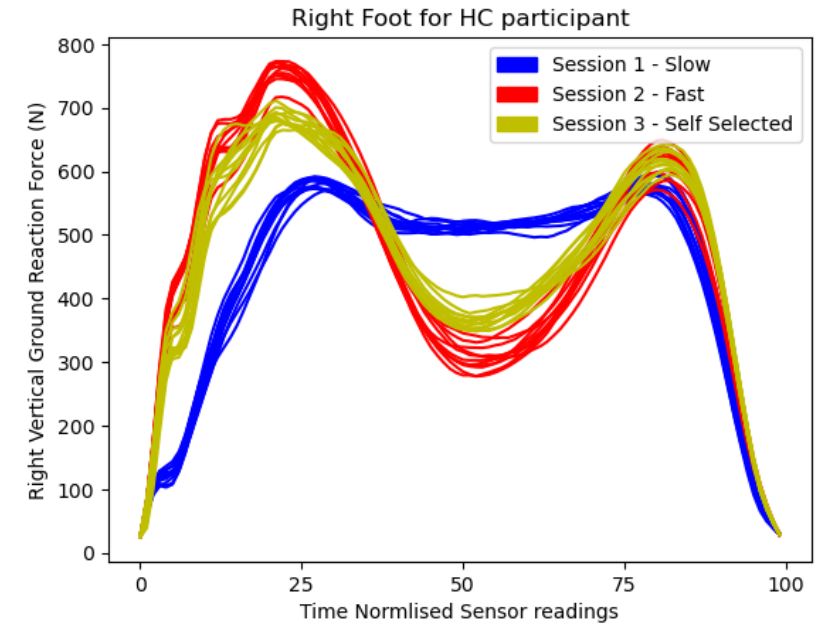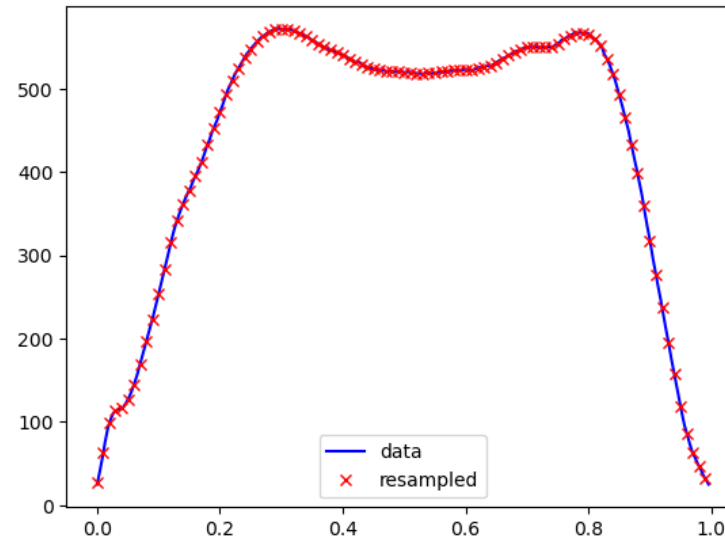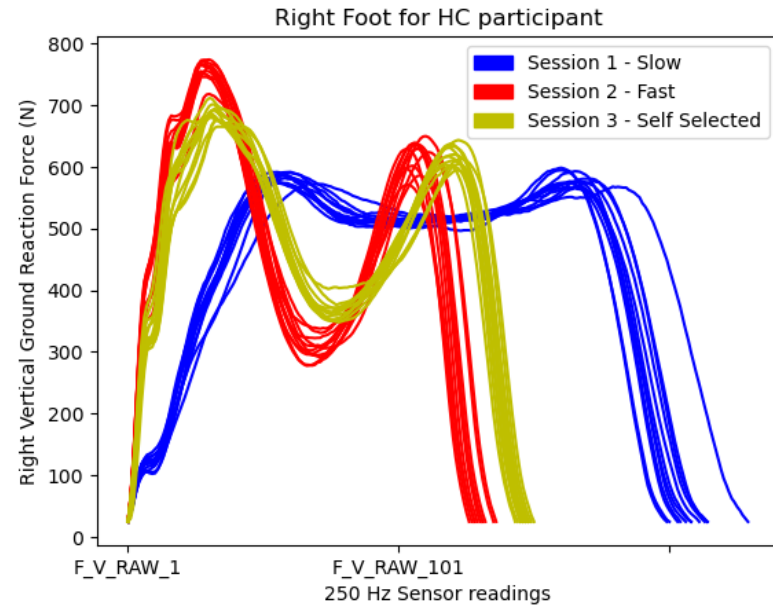Source: https://www.youtube.com/watch?v=7EdzFlRE8vY

Data Processing and Exploration

# Data Processing

90-400 readings

Re-sampling

100 readings
(100% stride)

The dataset is obtained from GaitRec, a large online database for gait analysis:
Horsak, B., Slijepcevic, D., Raberger, AM. et al. GaitRec, a large-scale ground reaction force dataset of healthy and impaired gait. Sci Data 7, 143 (2020)

All GRF readings are subsequently **normalized** by weight of participant

12 Selected Features

Vertical Force (Left and Right)

Vertical Force (Left – Right)

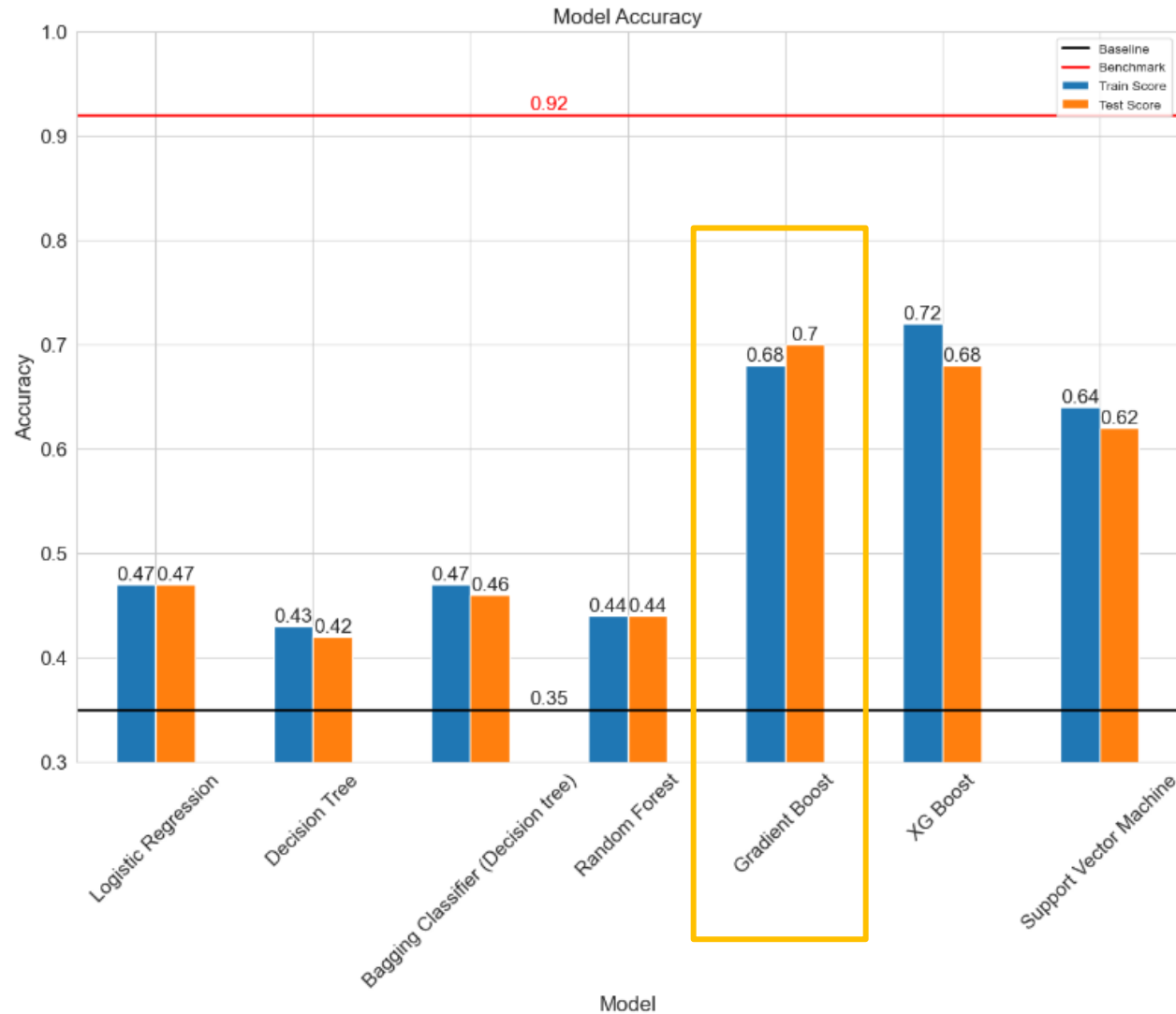Horizontal Force (Left - Right)

Age

BMI
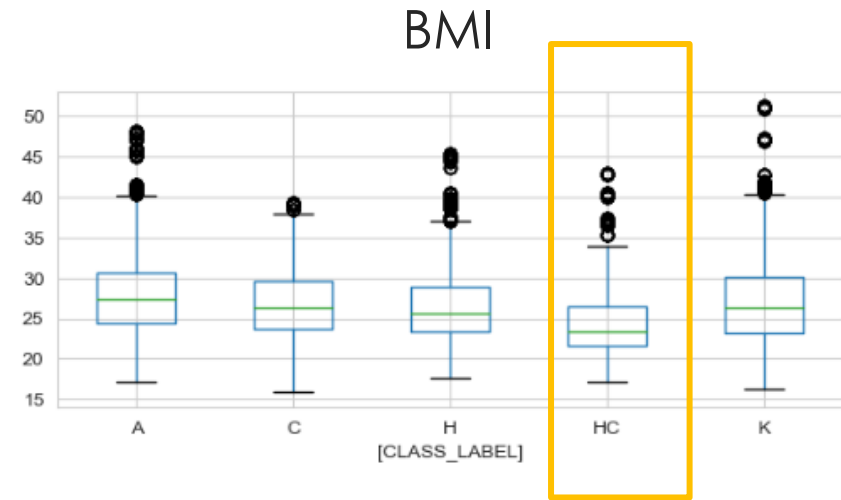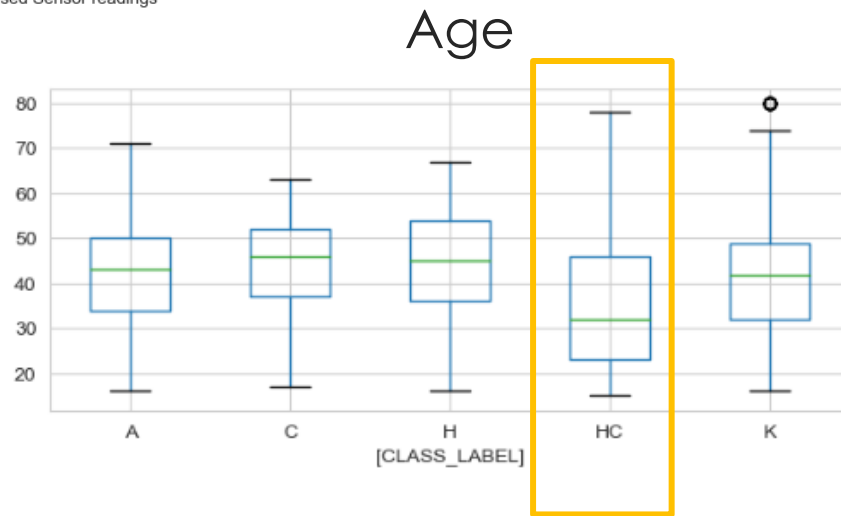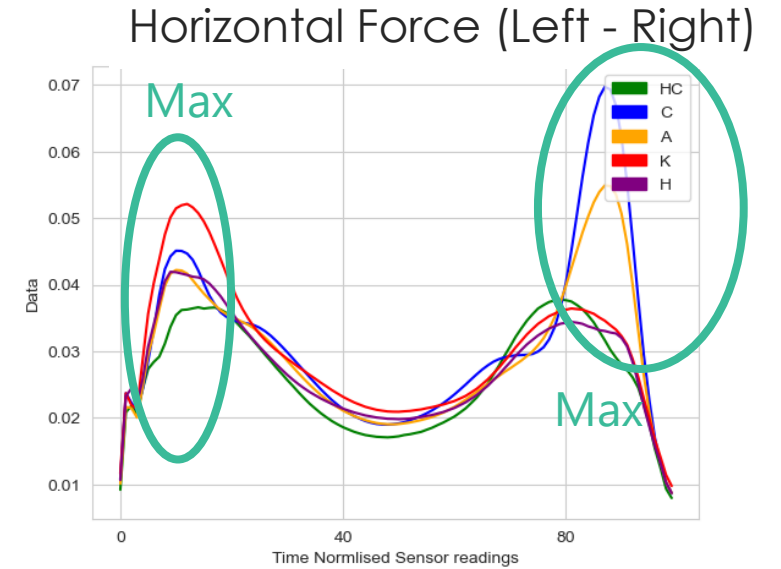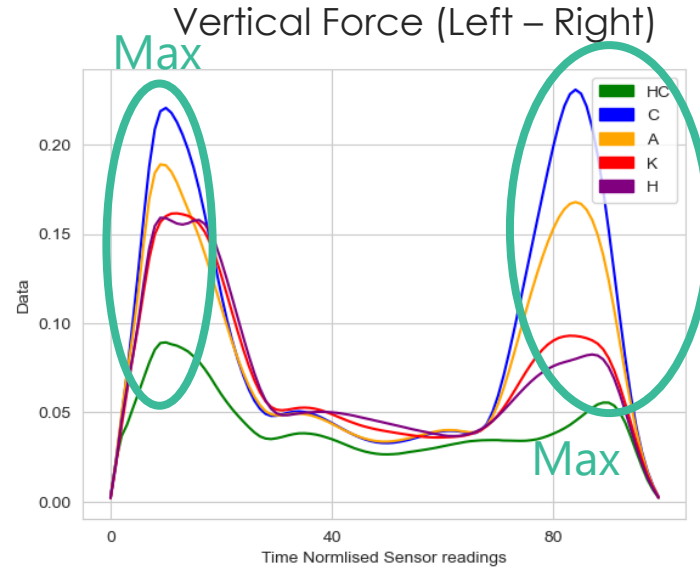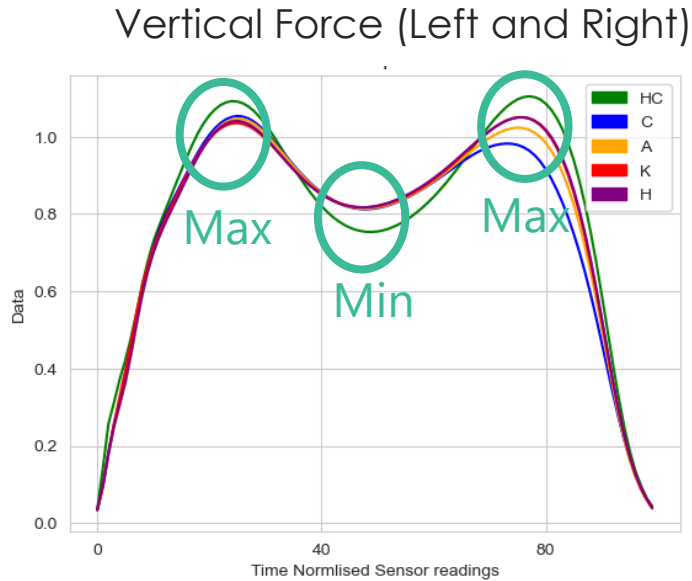
Healthy Control participants generally have lower Age and BMI

Model Evaluation and Conclusion

Model Evaluation

Gradient Boosted Classifier with accuracy of **70%** will be selected as the best alternative to the benchmark model of **92%** accuracy.

Benchmark: Neural Network with 1000 features

	precision	recall	f1-score	support
C	0.93	0.96	0.94	1904
K	0.94	0.91	0.92	2844
H	0.92	0.90	0.91	1635
A	0.90	0.95	0.92	3863
HC	0.95	0.77	0.85	660
accuracy			0.92	10906
macro avg	0.93	0.89	0.91	10906
weighted avg	0.92	0.92	0.92	10906

Confusion Matrix For NN with original 1000 Features
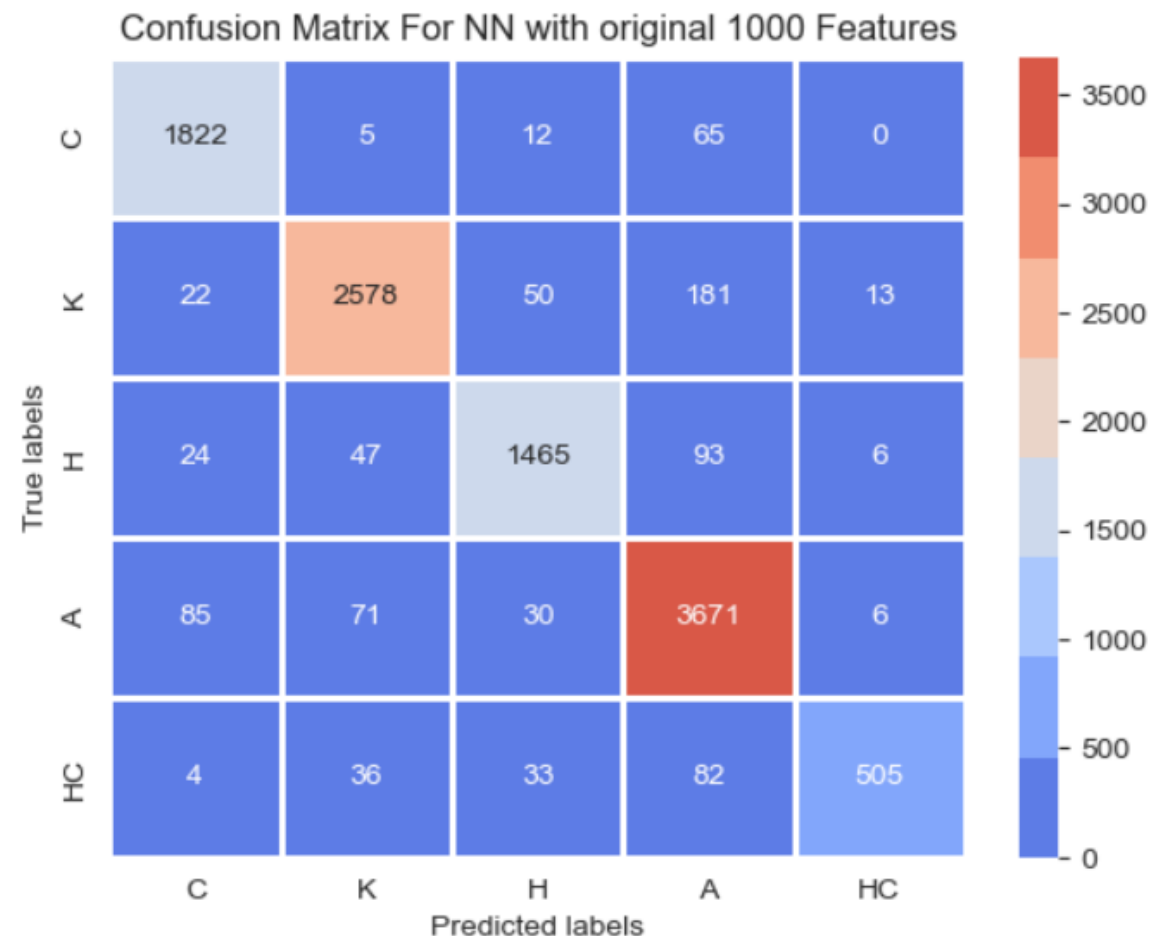
Gradient Boosted Trees with 12 selected features

	precision	recall	f1-score	support
A	0.71	0.77	0.74	3863
C	0.75	0.69	0.72	1904
H	0.71	0.64	0.67	1635
HC	0.61	0.54	0.57	660
K	0.69	0.69	0.69	2844
accuracy			0.70	10906
macro avg	0.69	0.67	0.68	10906
weighted avg	0.70	0.70	0.70	10906

Confusion Matrix for Gradient Boosting with Engineered Features
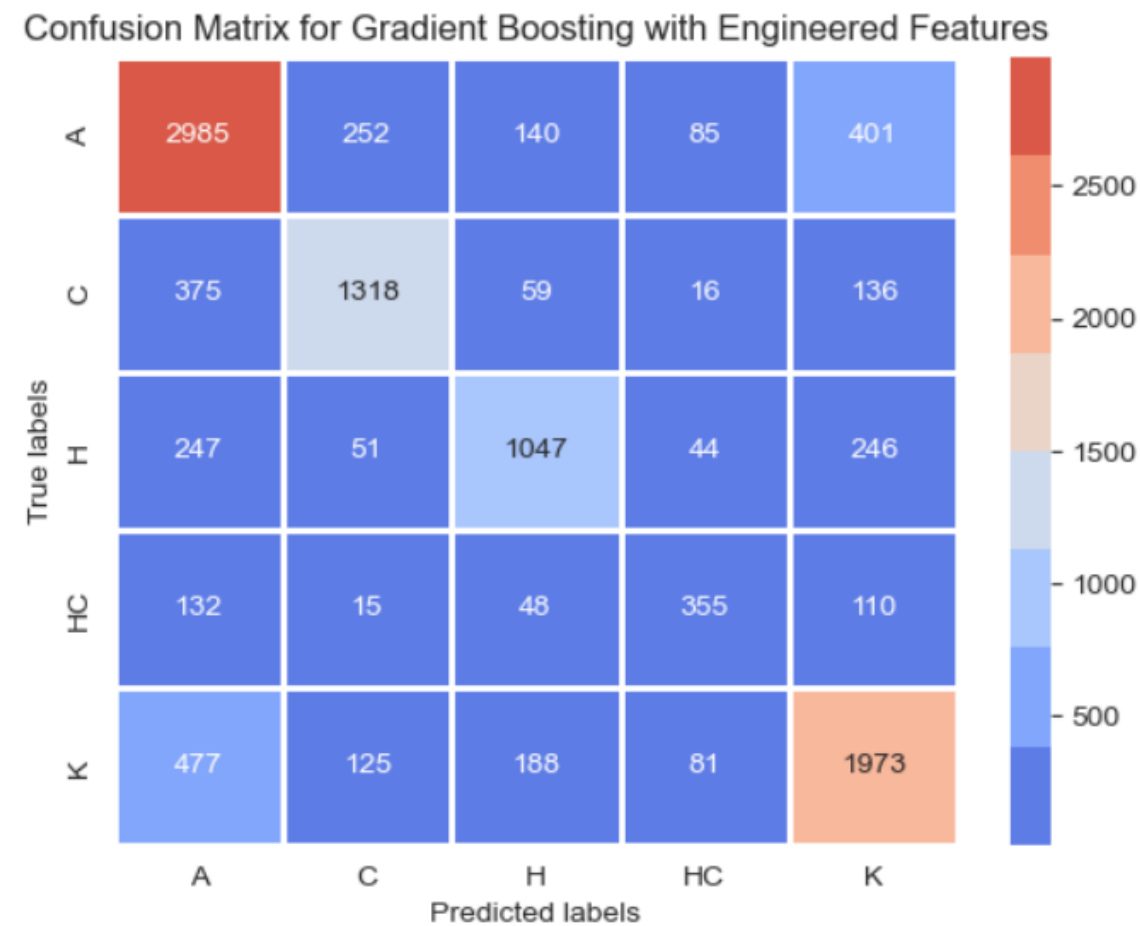
Conclusion

1. Using key features like BMI, Age, maximum landing force, push force, together with differences in vertical and horizontal forces between left and right leg can approximately reach 70% accuracy in classification of gait.

2. There is feasibility in exploring more key features and affordable sensors that target these key features in solutions and research available currently.
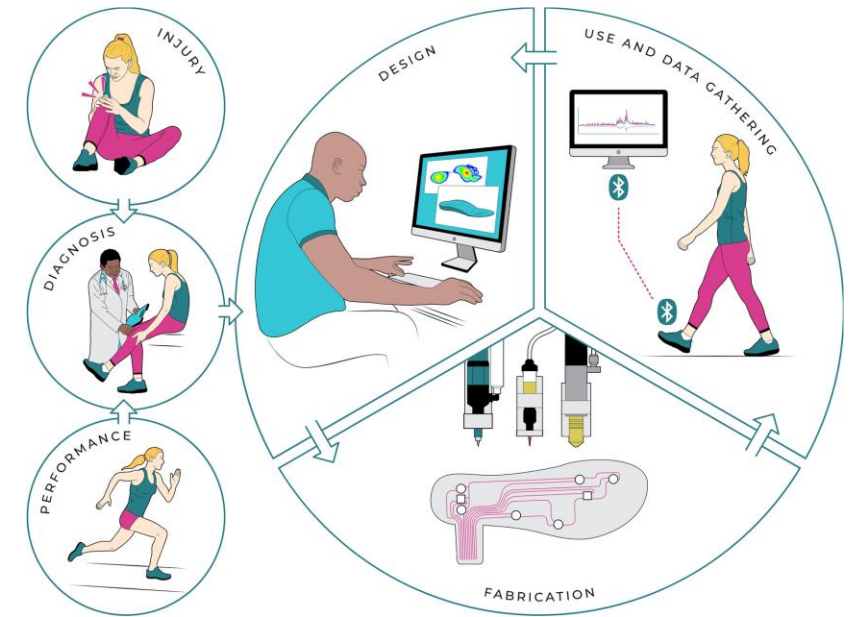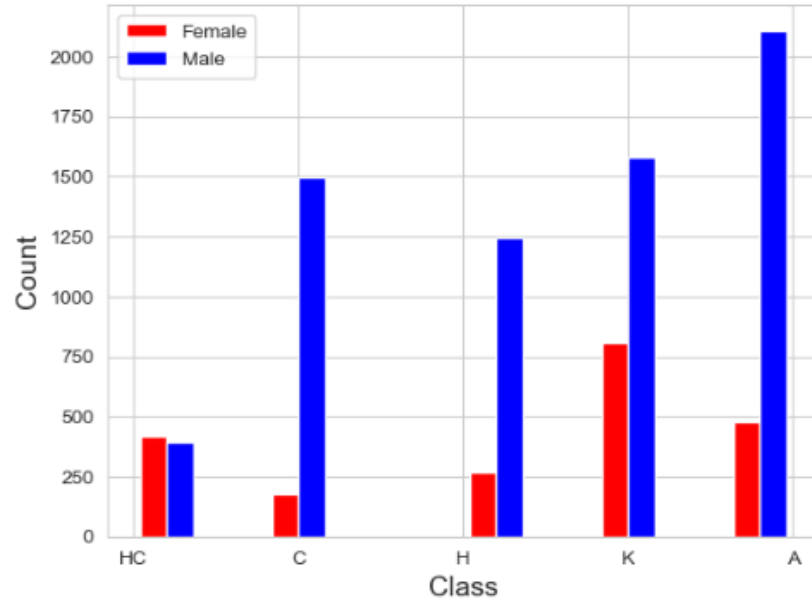

Source for illustration:
Binelli, M.R., van Dommelen, R., Nagel, Y. *et al.* Digital manufacturing of personalised footwear with embedded sensors. *Sci Rep* **13**, 1962 (2023). https://doi.org/10.1038/s41598-023-29261-0

Summary and Appendix

1. Prepare and process data

2. Feature Engineering

3. Feature Selection

4. Baseline model

5. Train and evaluate models

6. Compare with baseline and benchmark

7. Back to feature engineering

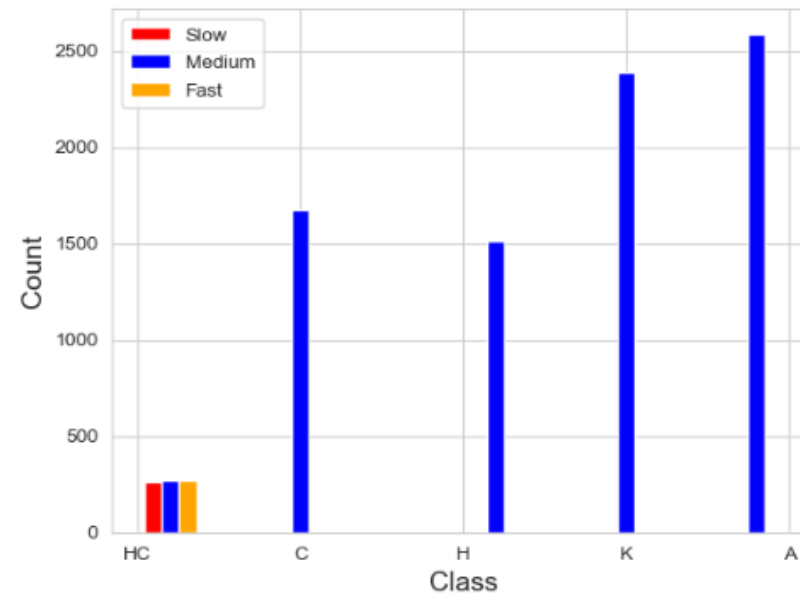8. Train and evaluate models

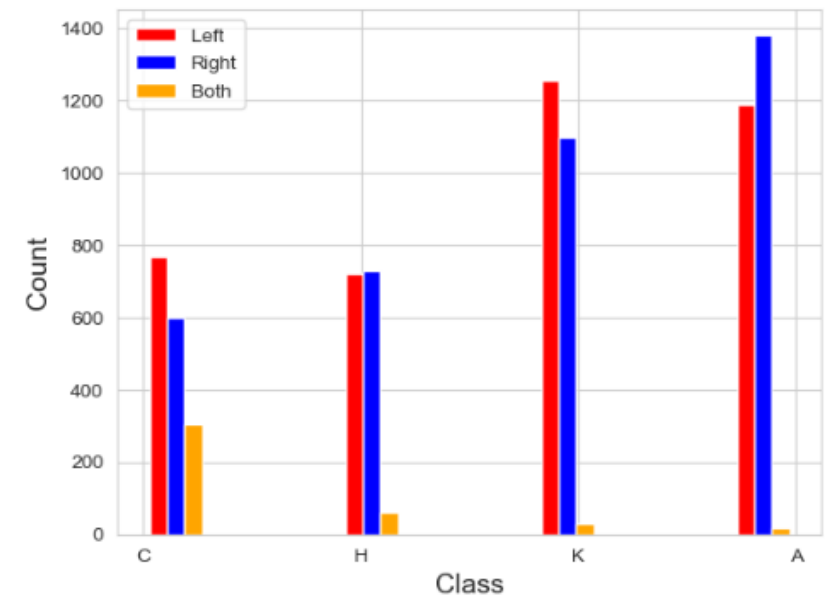9. Model Selection and Conclusion
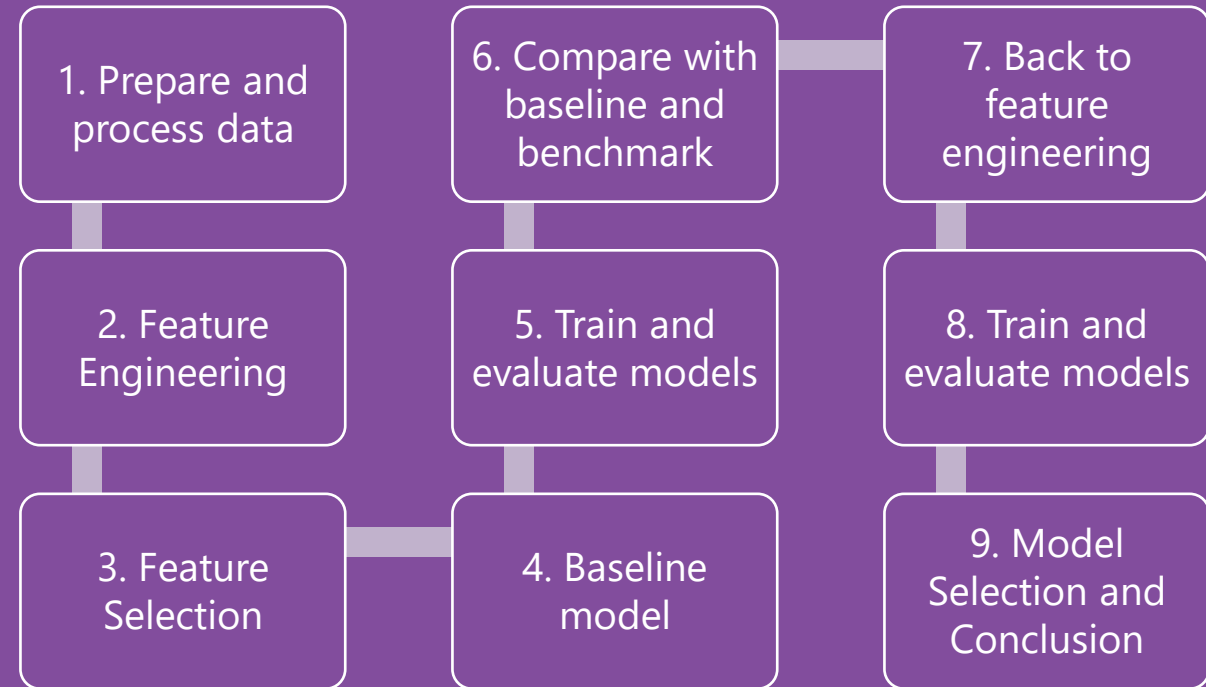
Distribution of classes
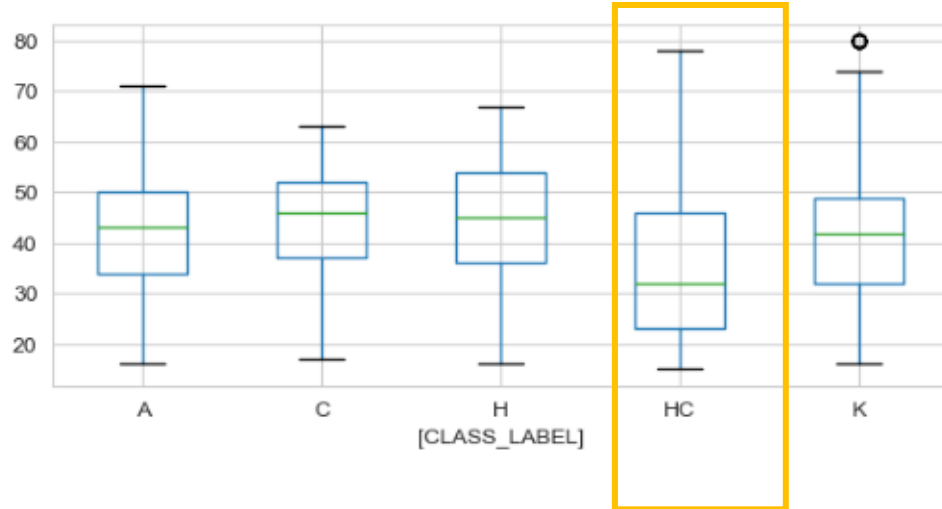

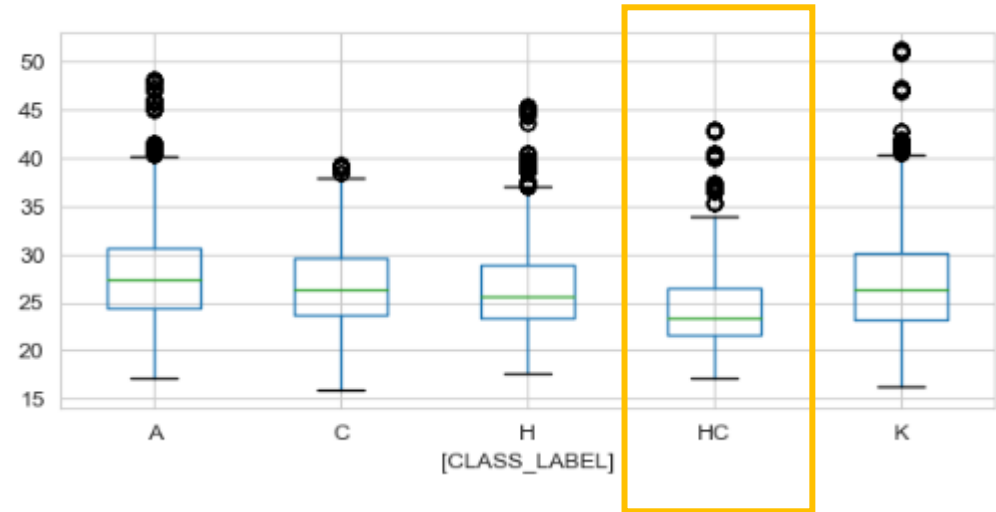Classes By Gender

Classes By Walking Speed

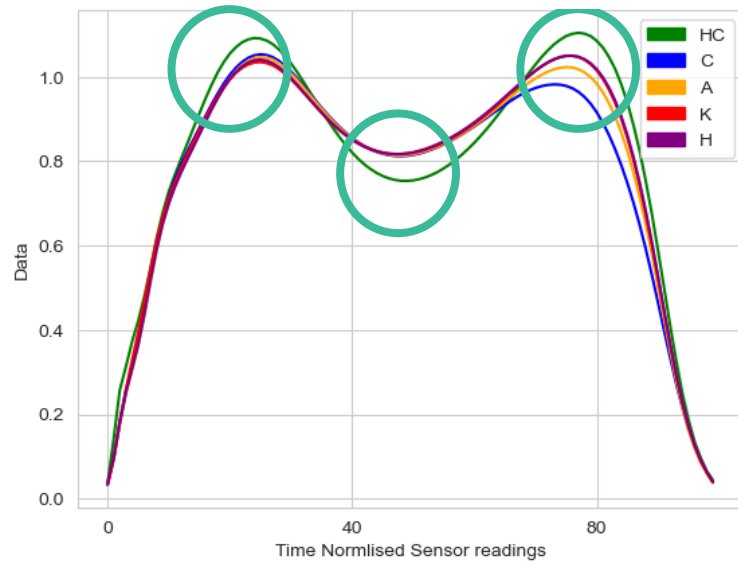Classes by Side

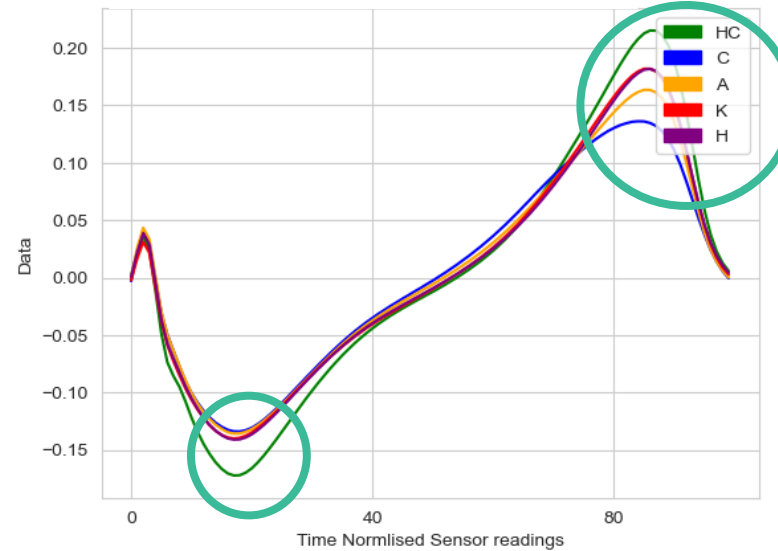# Age and BMI

AGE Boxplot

BMI Boxplot

Healthy Control participants generally have lower Age and BMI

Ground Reaction Forces Measurements
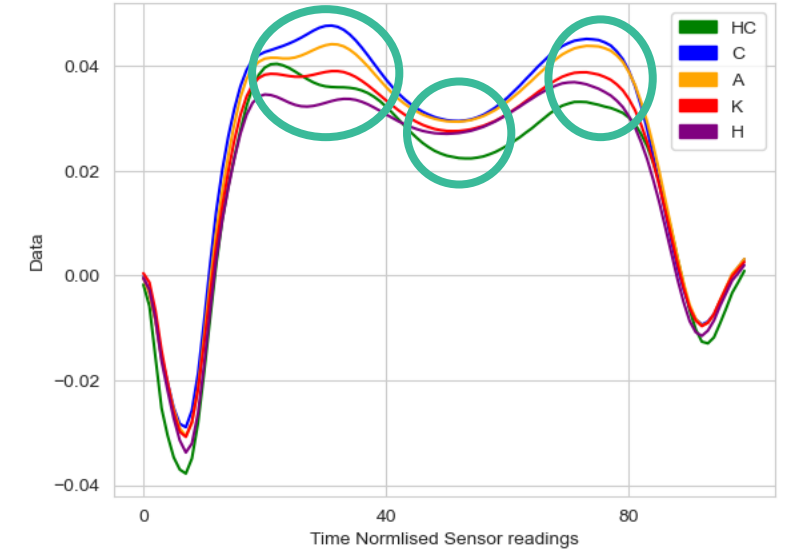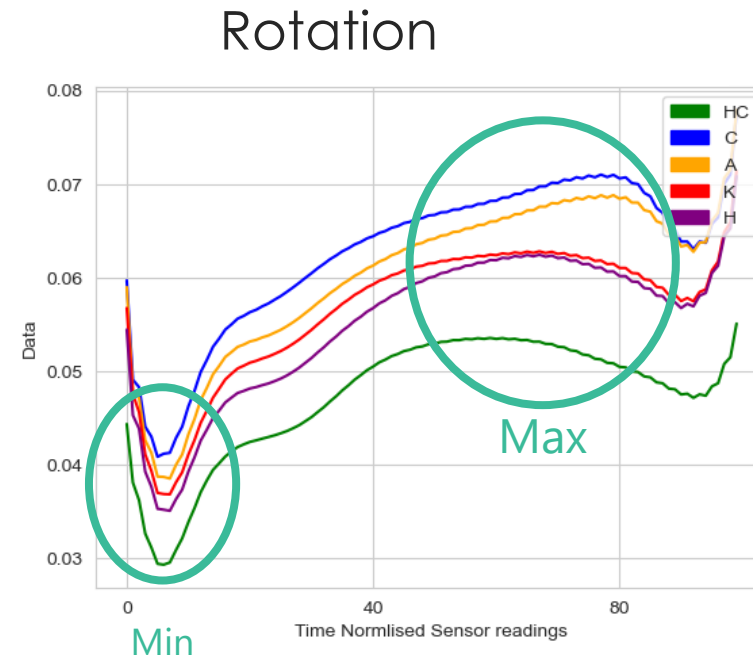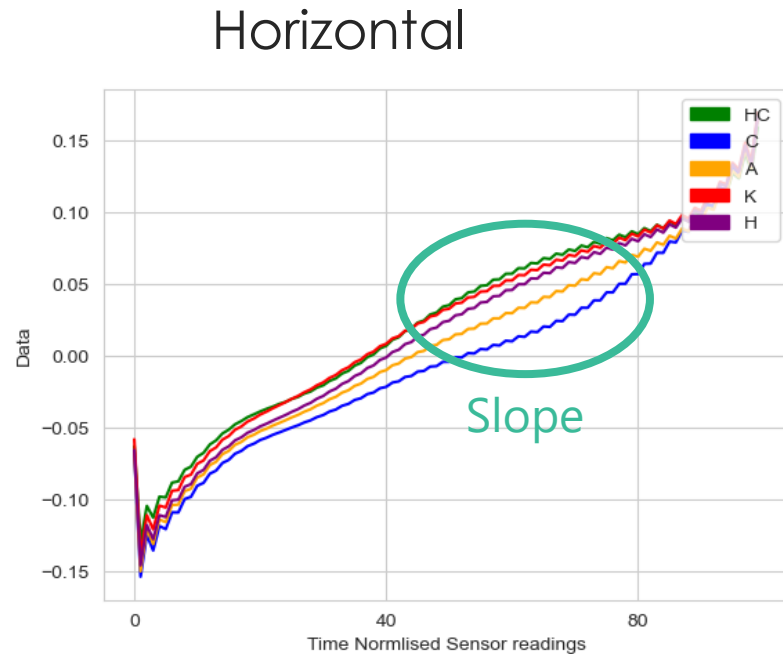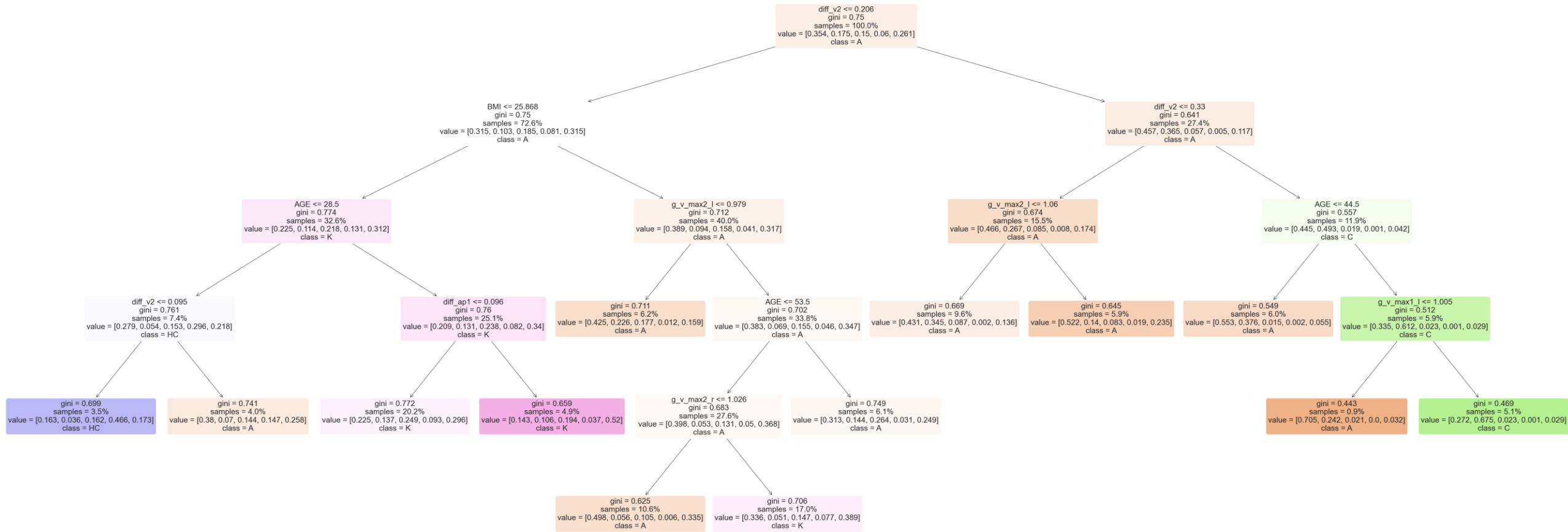

Vertical   Horizontal   Rotation

300 readings to **8** engineered features

Center of Pressure Measurements

200 readings to **3** engineered features

Decision Tree Classifier with 12 selected features

Training History of Neural Network

Accuracy For NN with original 1000 Features