Disease Classification Based on Gait Features

Dataset description – Gait in Neurodegenerative Disease Database v1.0.0

Taken from physionet

Records in this database are from patients with Parkinson's disease (n = 15), Huntington's disease (n = 20), or amyotrophic lateral sclerosis (n = 13). Records from 16 healthy control subjects are also included here.

Method used while collecting data –

Force-Sensitive Resistors (FSRs):

- These are sensors that change their resistance based on the force or pressure applied.
- They were placed under the foot (possibly in insoles).
- The output is electrical (voltage or current) and is roughly proportional to the applied force (i.e., the pressure of the foot on the ground).

Raw Data:

- The data recorded from these sensors is in the form of time-series signals, where the values rise and fall depending on whether the foot is in contact with the ground.
- For example, when the foot touches the ground, the force increases → sensor output goes up.
- When the foot lifts off, the force drops → sensor output goes down.

Stride-to-Stride Measures:

- A stride is one complete cycle of movement by a foot (from contact → lift-off → next contact).
- From the force signal, researchers can determine:
    - Contact time: How long each foot was in contact with the ground during each stride.
    - Timing between steps, variability, asymmetry, etc.

Derived Measures:

- These contact times are not directly measured—they're calculated ("derived") from the sensor data using a threshold or signal processing techniques.
- For example, when the signal crosses a certain force threshold, that indicates foot contact or lift-off.

The diseases are – Parkinson, Amyotrophic Lateral Sclerosis, huntington's, control

Files are –

Suffix	Contents
.hea	Header (text)
.let	Left foot signal (binary)
.rit	Right foot signal (binary)
.ts	Derived time series (text)

In .ts files the contents are -

Column	Contents
1	Elapsed Time (sec)
2	Left Stride Interval (sec)
3	Right Stride Interval (sec)
4	Left Swing Interval (sec)
5	Right Swing Interval (sec)
6	Left Swing Interval (% of stride)
7	Right Swing Interval (% of stride)
8	Left Stance Interval (sec)
9	Right Stance Interval (sec)
10	Left Stance Interval (% of stride)
11	Right Stance Interval (% of stride)
12	Double Support Interval (sec)
13	Double Support Interval (% of stride)

Stride - one full cycle of walking — from when one foot hits the ground to the next time that same foot hits the ground again.

Stance - The stance phase is the part of the stride when the foot is on the ground. Ideally 60% of gait cycle

Swing - The swing phase is when the foot is in the air, moving forward to take the next step. Ideally 40%

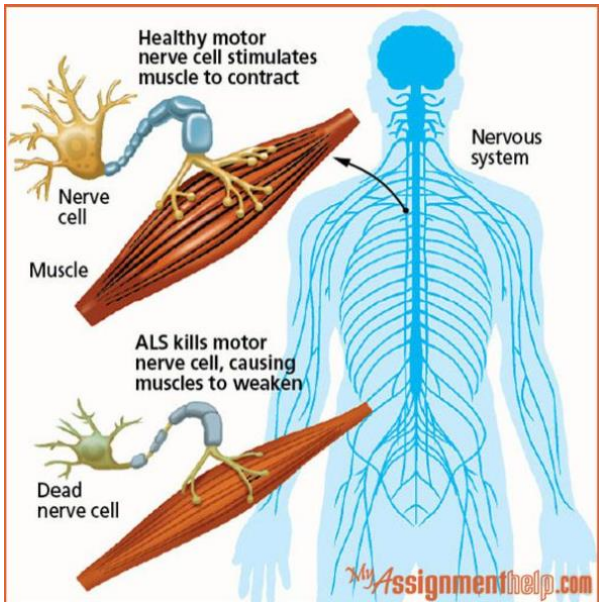
Calculation of gait asymmetry and its role in determining the disease –

Parameter	What It Shows	Healthy Gait Expectation

Stride asymmetry	Difference in length/duration of strides between left and right leg	< 5% difference
Swing asymmetry	Time spent in air (left vs right)	Balanced swing duration
Stance asymmetry	Time spent on ground per foot	Equal weight-bearing

Diseases in the dataset are –

ALS (Amyotrophic Lateral Sclerosis) – Affects motor neurons, leading to muscle weakness and paralysis.



- Destroys both upper motor neurons (brain to spinal cord) and lower motor neurons (spinal cord to muscles)
- Muscles work because motor neurons send electrical signals to them. These signals cause the muscle fibers to contract.
- In normal person, the activity is predictable recorded using EMG. But is abnormal for person having ALS.
- Brain sends signals through motor neurons to muscles. But since motor neurons destroyed, electrical signals received by muscles are not smooth. Irregularly muscles might move trying to compensate.
- Swing is lower, since difficulty in lifting foot while walking. Stance is longer than regular since more time on ground. Muscles fail to contract. Which means they cannot create force or make movement.
- Overall stride is less.

Parkinson’s disease –

degeneration of nerve cells in the substantia nigra, a midbrain region that provides dopamine to the basal ganglia, a system involved in voluntary motor control.

Shuffling steps, small stride length.

Reduced arm swing, especially asymmetrically.

Festination – walking starts slow and speeds up uncontrollably.

Difficulty initiating movement (“freezing of gait”).

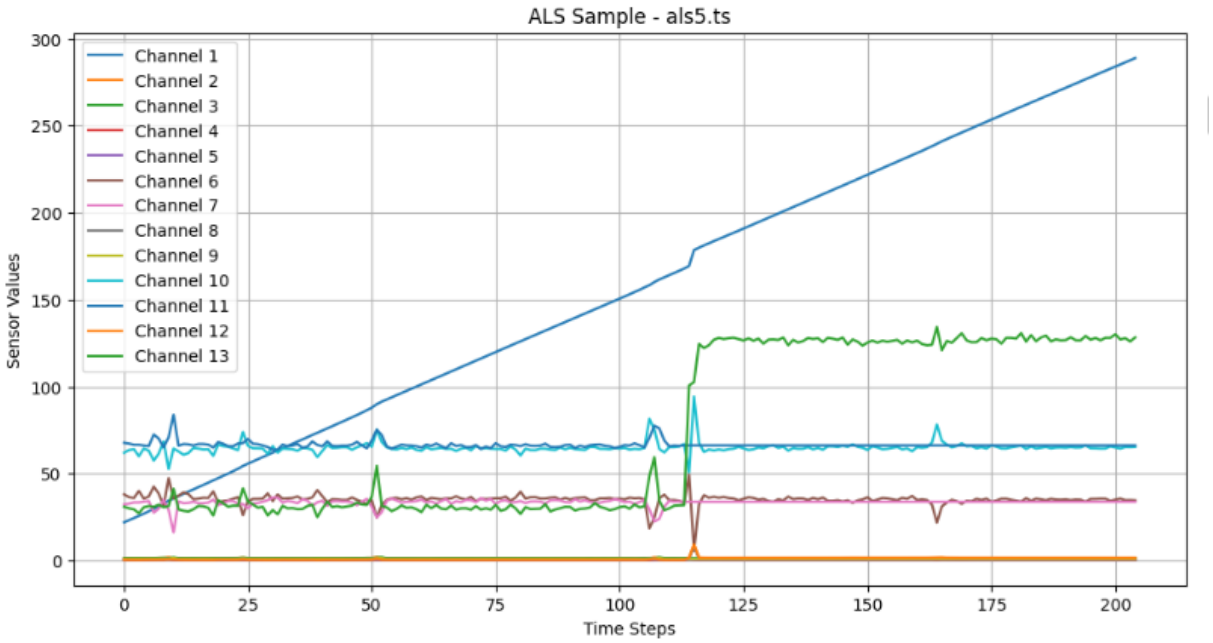
Stooped posture, less trunk rotation.

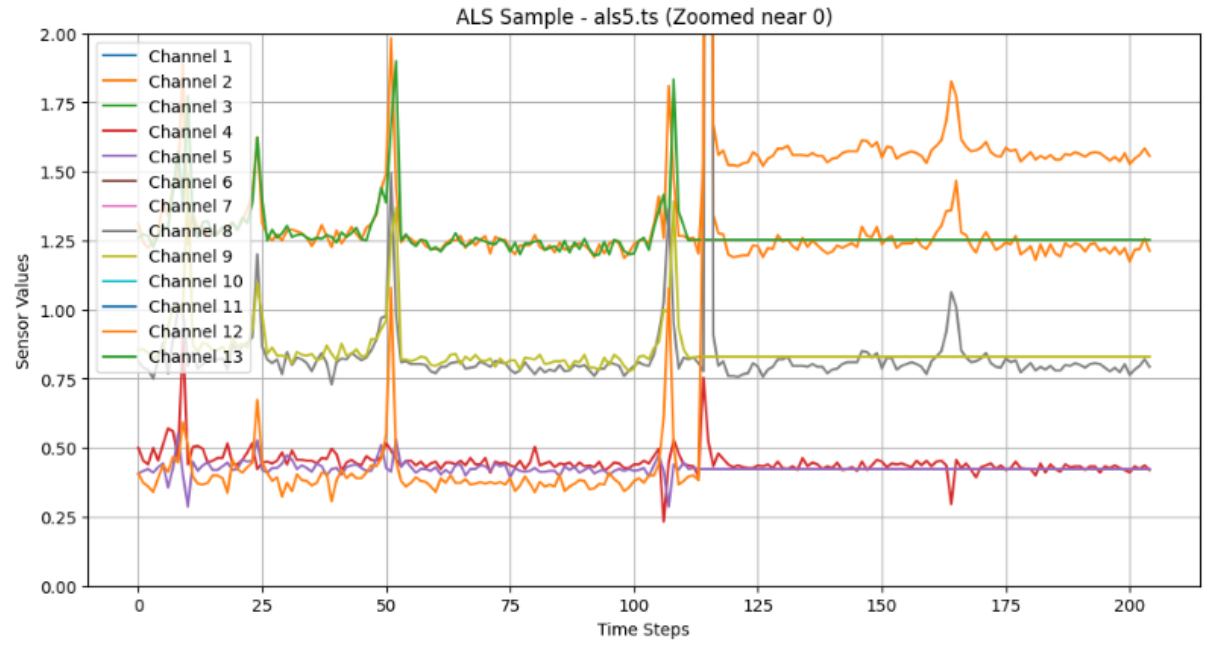
Huntington’s disease -

The movement disorders related to Huntington's disease may cause movements that can't be controlled, called chorea. Chorea are involuntary movements affecting all the muscles of the body, specifically the arms and legs, the face and the tongue. They also can affect the ability to make voluntary movements. Symptoms may include:

- Involuntary jerking or writhing movements.
- Muscle rigidity or muscle contracture.
- Slow or unusual eye movements.
- Trouble walking or keeping posture and balance.

111	165.4000	1.2567	1.2500	0.4367	0.4233	34.75	33.87	0.8200	0.8267	65.25	66.13	0.3967	31.56
112	166.6600	1.2600	1.2533	0.4333	0.4267	34.39	34.04	0.8267	0.8267	65.61	65.96	0.4000	31.75
113	167.8633	1.2033	1.2533	0.3967	0.4233	32.96	33.78	0.8067	0.8300	67.04	66.22	0.3833	31.86
114	169.3933	1.5300	1.2533	0.7533	0.4233	49.24	33.78	0.7767	0.8300	50.76	66.22	1.5400	100.65
115	178.5767	9.1833	1.2533	0.5233	0.4233	5.70	33.78	8.6600	0.8300	94.30	66.22	9.4233	102.61
116	179.9200	1.3433	1.2533	0.4333	0.4233	32.26	33.78	0.9100	0.8300	67.74	66.22	1.6733	124.57
117	181.1967	1.2767	1.2533	0.4800	0.4233	37.60	33.78	0.7967	0.8300	62.40	66.22	1.5600	122.19
118	182.4700	1.2733	1.2533	0.4600	0.4233	36.13	33.78	0.8133	0.8300	63.87	66.22	1.5767	123.82
119	183.6700	1.2000	1.2533	0.4400	0.4233	36.67	33.78	0.7600	0.8300	63.33	66.22	1.5233	126.94
120	184.8600	1.1900	1.2533	0.4300	0.4233	36.13	33.78	0.7600	0.8300	63.87	66.22	1.5233	128.01
121	186.0533	1.1933	1.2533	0.4367	0.4233	36.59	33.78	0.7567	0.8300	63.41	66.22	1.5200	127.37
122	187.2500	1.1967	1.2533	0.4300	0.4233	35.93	33.78	0.7667	0.8300	64.07	66.22	1.5300	127.86
123	188.4467	1.1967	1.2533	0.4267	0.4233	35.65	33.78	0.7700	0.8300	64.35	66.22	1.5333	128.13
124	189.6800	1.2333	1.2533	0.4267	0.4233	34.59	33.78	0.8067	0.8300	65.41	66.22	1.5700	127.30
125	190.9100	1.2300	1.2533	0.4367	0.4233	35.50	33.78	0.7933	0.8300	64.50	66.22	1.5567	126.56
126	192.1000	1.1900	1.2533	0.4333	0.4233	36.41	33.78	0.7567	0.8300	63.59	66.22	1.5200	127.73
127	193.3233	1.2233	1.2533	0.4400	0.4233	35.97	33.78	0.7833	0.8300	64.03	66.22	1.5467	126.43
128	194.5433	1.2200	1.2533	0.4267	0.4233	34.97	33.78	0.7933	0.8300	65.03	66.22	1.5567	127.60
129	195.8133	1.2700	1.2533	0.4500	0.4233	35.43	33.78	0.8200	0.8300	64.57	66.22	1.5833	124.67
130	197.0600	1.2467	1.2533	0.4267	0.4233	34.22	33.78	0.8200	0.8300	65.78	66.22	1.5833	127.01
131	198.3133	1.2533	1.2533	0.4233	0.4233	33.78	33.78	0.8300	0.8300	66.22	66.22	1.5933	127.13
132	199.5433	1.2300	1.2533	0.4333	0.4233	35.23	33.78	0.7967	0.8300	64.77	66.22	1.5600	126.83





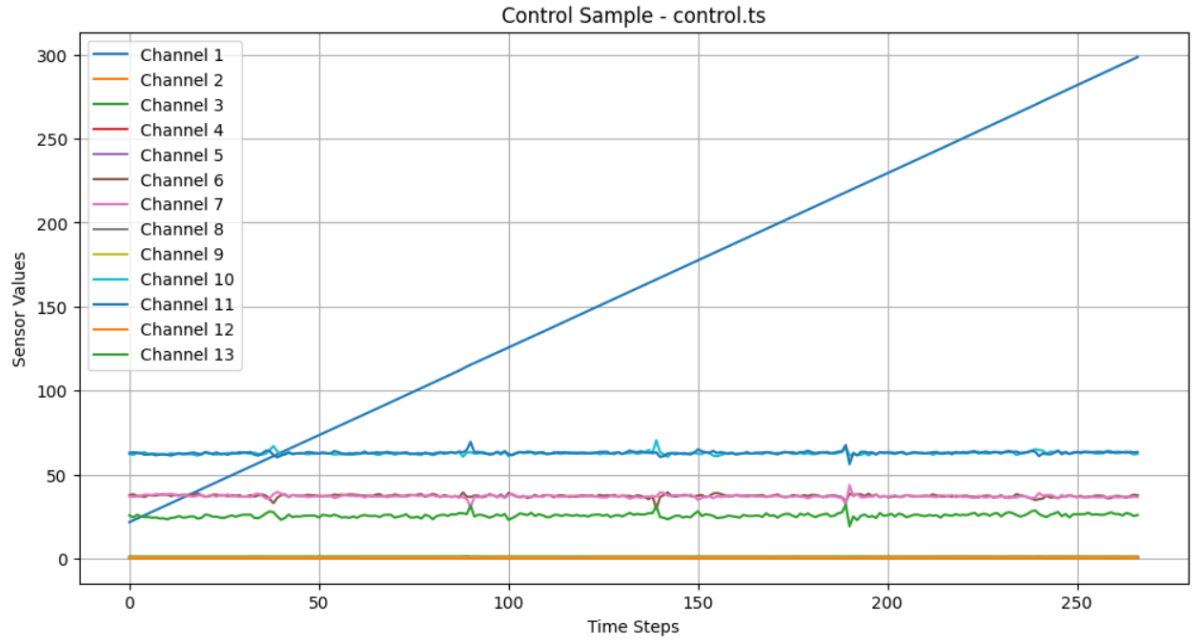
The above 2 graphs show stance, swing and stride characteristics of a person with ALS.

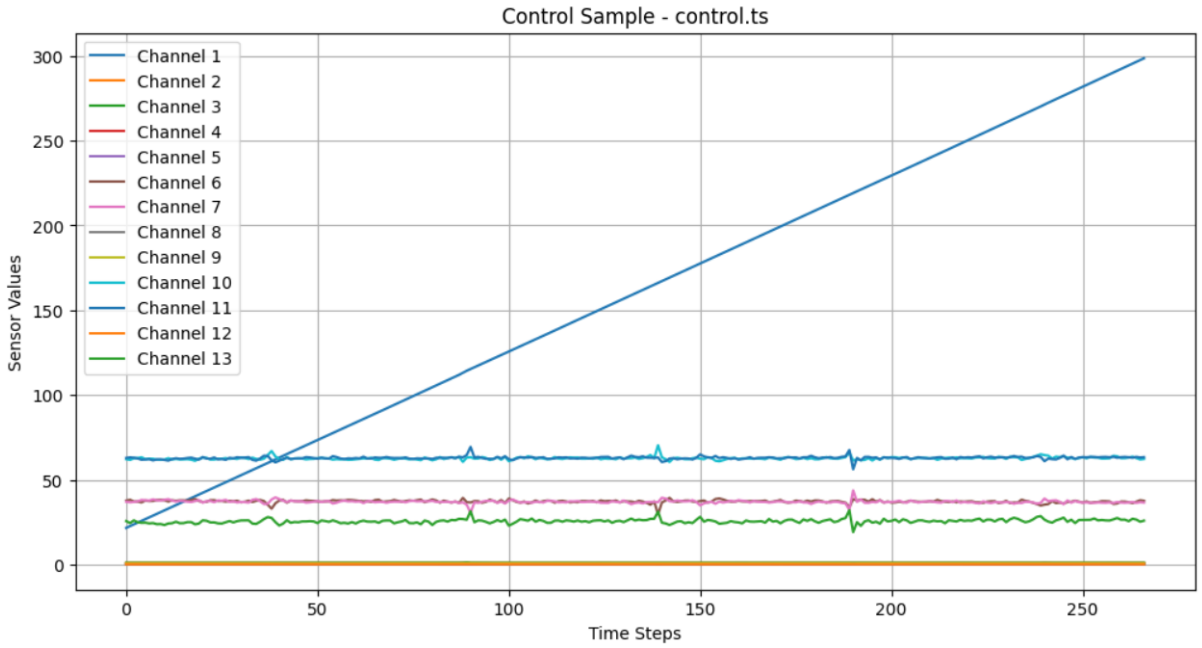
Firstly, the line (channel 1) extending from sensor value – 25 to 270, indicates the recorded elapsed time in seconds. Which means, the start time of a gait cycle. It is linearly increasing indicating constant time for every gait cycle. However, at around 112 time step, an increase in gait cycle time is recorded.

At the same point, there is an abrupt increase in double support interval from 0.35 to 1.5 sec. Showing that both the foot were placed for a longer time. This indicates balance issues where the person is unable to move forward by raising his legs.

Over all in the stride and stance time intervals showed sudden spikes indicating longer times of foot on ground at certain time steps showing the difficulty in walking.

For a normal person -



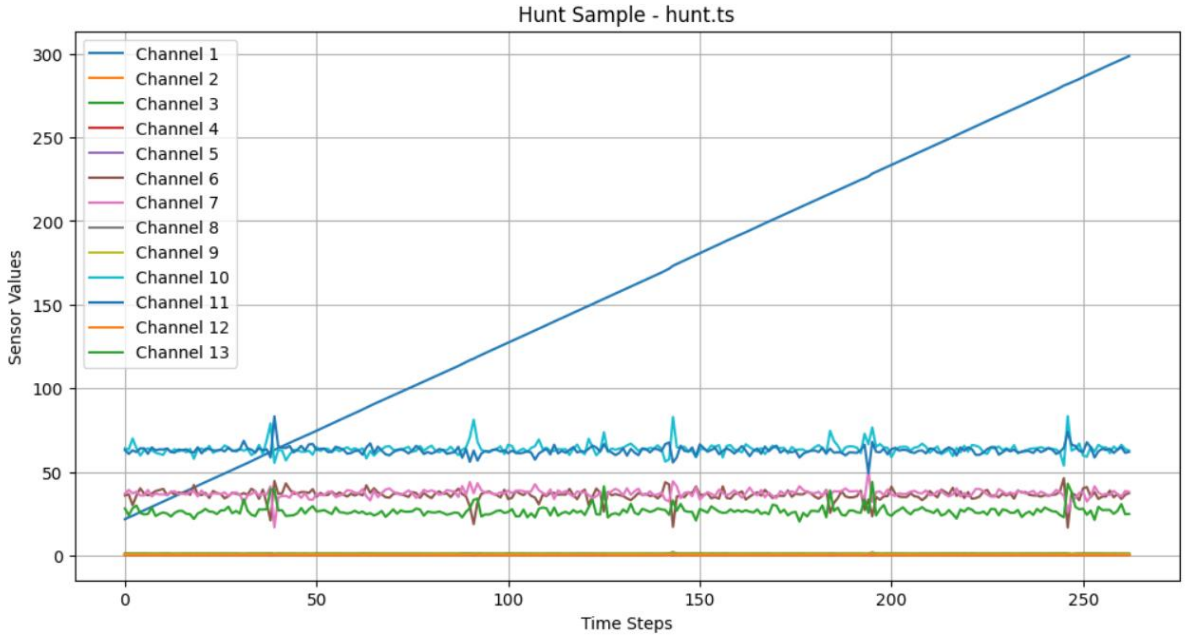


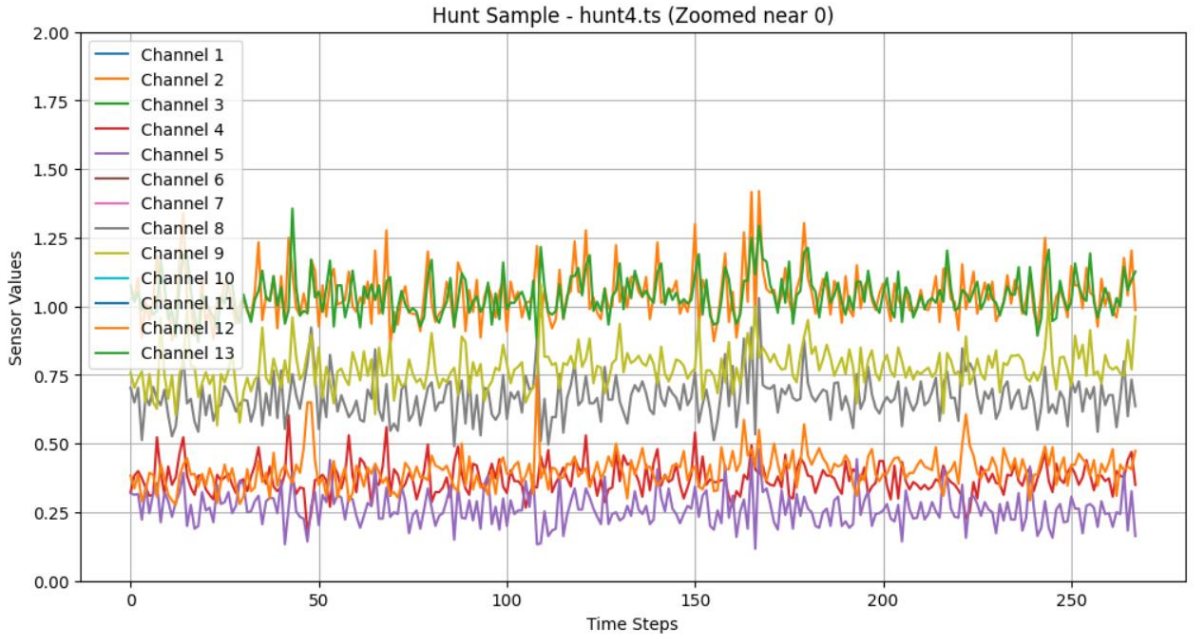
The above two graphs are of a normal person's walking characteristics. A steady increase in Channel 1 (elapsed time) shows that there is a constant time interval in every gait cycle.

Channel 10 and 11 indicate percentage of stance in the gait cycle. They are close to 60% which are ideal measures.

Channel 6 and 7 11 indicate percentage of swing in the gait cycle. They are close to 40% which are ideal measures.

Viewing the stance, stride and swing intervals, they do not have sudden abrupt increase in the intervals, which indicate stability in movement.





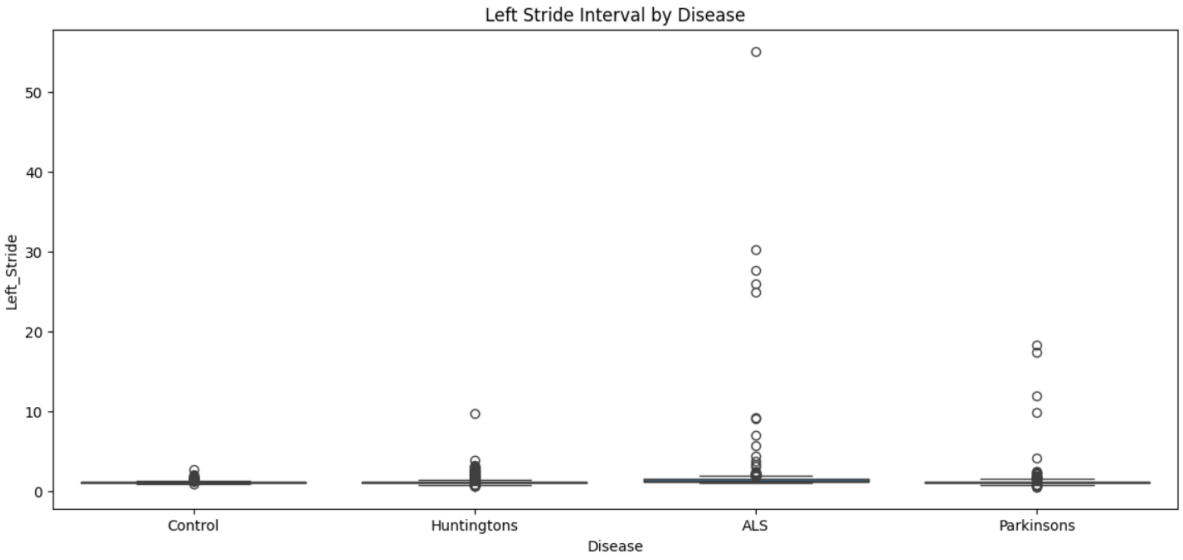
Above 2 graphs show gait characteristics of a person with huntingons disease. Because a person with Huntington's is effected with chorea, which is caused due to damage to basal ganglia – the part of brain that controls voluntary movements, his movements are noticed to be irregular.

Looking at the 2nd graph, which shows stance, swing and stride intervals has very jerky patterns. Which means his movements aren't constant. Some have longer intervals of foot, others have shorter.

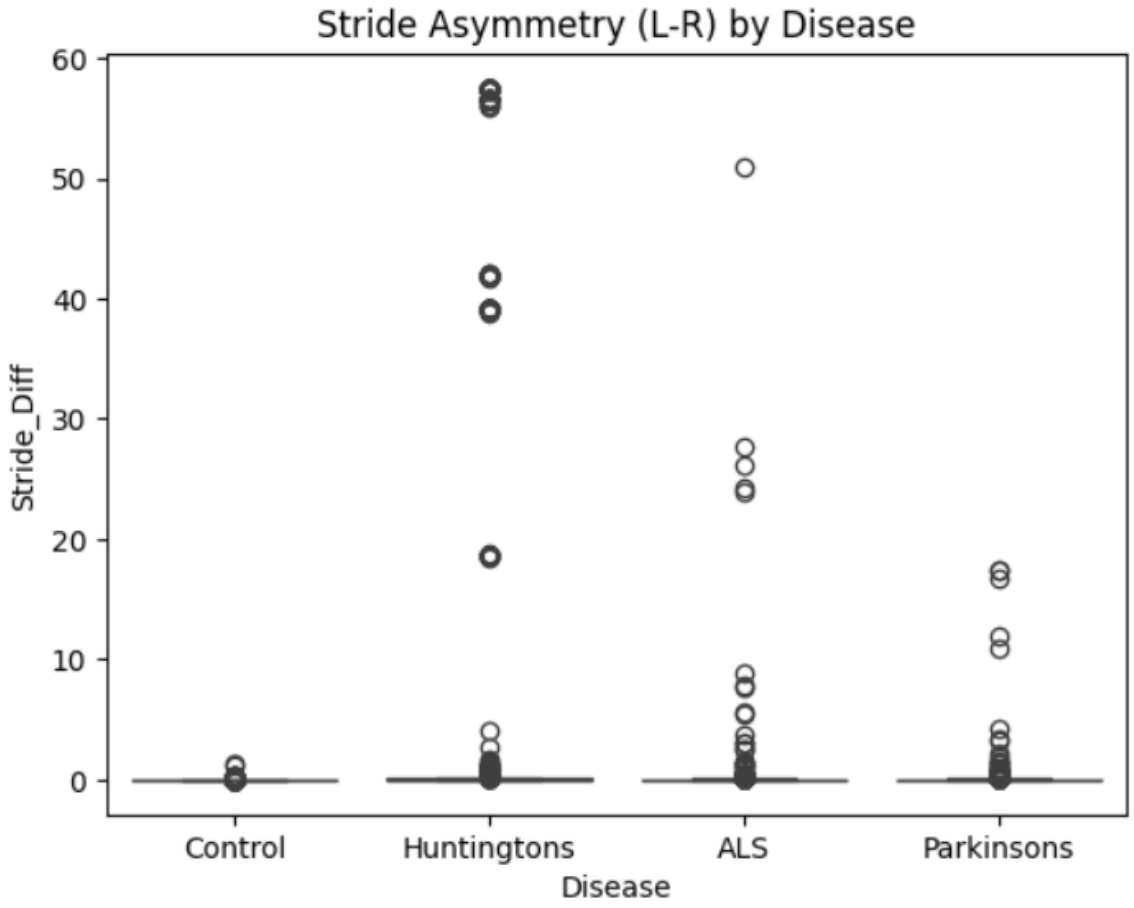
	0	1	2	3	4	5	6	7	8	9	10	11	12
0	22.4867	1.0500	1.0300	0.3967	0.3867	37.78	37.54	0.6533	0.6433	62.22	62.46	0.2667	25.40
1	23.5467	1.0600	1.0500	0.3967	0.3933	37.42	37.46	0.6633	0.6567	62.58	62.54	0.2700	25.47
2	24.5700	1.0233	1.0467	0.3833	0.3733	37.46	35.67	0.6400	0.6733	62.54	64.33	0.2667	26.06
3	25.6233	1.0533	1.0467	0.4067	0.3967	38.61	37.90	0.6467	0.6500	61.39	62.10	0.2500	23.73
4	26.6833	1.0600	1.0500	0.4067	0.4067	38.36	38.73	0.6533	0.6433	61.64	61.27	0.2467	23.27
5	27.7400	1.0567	1.0567	0.3933	0.3967	37.22	37.54	0.6633	0.6600	62.78	62.46	0.2667	25.24
6	28.7867	1.0467	1.0633	0.3933	0.4067	37.58	38.24	0.6533	0.6567	62.42	61.76	0.2467	23.57
7	29.8367	1.0500	1.0400	0.3967	0.3967	37.78	38.14	0.6533	0.6433	62.22	61.86	0.2567	24.44
8	30.9033	1.0667	1.0533	0.4100	0.3933	38.44	37.34	0.6567	0.6600	61.56	62.66	0.2633	24.69
9	31.9500	1.0467	1.0733	0.3933	0.4067	37.58	37.89	0.6533	0.6667	62.42	62.11	0.2467	23.57
10	33.0233	1.0733	1.0567	0.3933	0.4100	36.65	38.80	0.6800	0.6467	63.35	61.20	0.2700	25.16
11	34.0767	1.0533	1.0567	0.3967	0.3933	37.66	37.22	0.6567	0.6633	62.34	62.78	0.2633	25.00
12	35.1433	1.0667	1.0467	0.4167	0.3967	39.06	37.90	0.6500	0.6500	60.94	62.10	0.2533	23.75
13	36.2033	1.0600	1.0567	0.4133	0.3867	38.99	36.59	0.6467	0.6700	61.01	63.41	0.2600	24.53
14	37.2300	1.0267	1.0467	0.3900	0.3733	37.99	35.67	0.6367	0.6733	62.01	64.33	0.2633	25.65
15	38.2700	1.0400	1.0433	0.4000	0.3900	38.46	37.38	0.6400	0.6533	61.54	62.62	0.2500	24.04
16	39.3233	1.0533	1.0533	0.4000	0.3933	37.97	37.34	0.6533	0.6600	62.03	62.66	0.2600	24.68
17	40.3567	1.0333	1.0367	0.3967	0.3900	38.39	37.62	0.6367	0.6467	61.61	62.38	0.2467	23.87
18	41.3800	1.0233	1.0367	0.3800	0.3867	37.13	37.30	0.6433	0.6500	62.87	62.70	0.2567	25.08
19	42.4167	1.0367	1.0300	0.3867	0.3833	37.30	37.22	0.6500	0.6467	62.70	62.78	0.2667	25.72
20	43.5000	1.0833	1.0433	0.4267	0.4000	39.38	38.34	0.6567	0.6433	60.62	61.66	0.2567	23.69
21	44.5200	1.0200	1.0567	0.3900	0.3900	38.24	36.91	0.6300	0.6667	61.76	63.09	0.2400	23.53
22	45.6300	1.1100	1.0600	0.4300	0.4167	38.74	39.31	0.6800	0.6433	61.26	60.69	0.2633	23.72
23	46.7367	1.1067	1.1267	0.4067	0.4467	36.75	39.64	0.7000	0.6800	63.25	60.36	0.2533	22.89
24	47.8000	1.0633	1.0700	0.4033	0.4000	37.93	37.38	0.6600	0.6700	62.07	62.62	0.2600	24.45
25	48.8333	1.0333	1.0500	0.3900	0.3933	37.74	37.46	0.6433	0.6567	62.26	62.54	0.2500	24.19
26	49.9233	1.0900	1.0633	0.4167	0.4000	38.23	37.62	0.6733	0.6633	61.77	62.38	0.2733	25.08

Below are the seaborn Box plots of 64 subjects at all intervals.

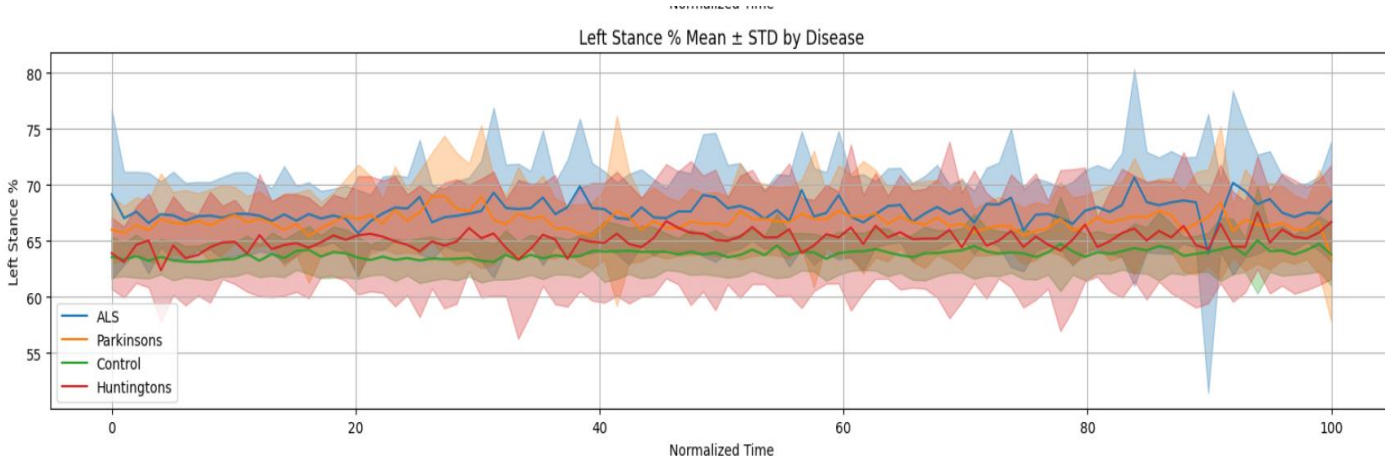
Disease	Left_Stride
ALS	2550
Control	4076
Huntingtons	4846
Parkinsons	3688



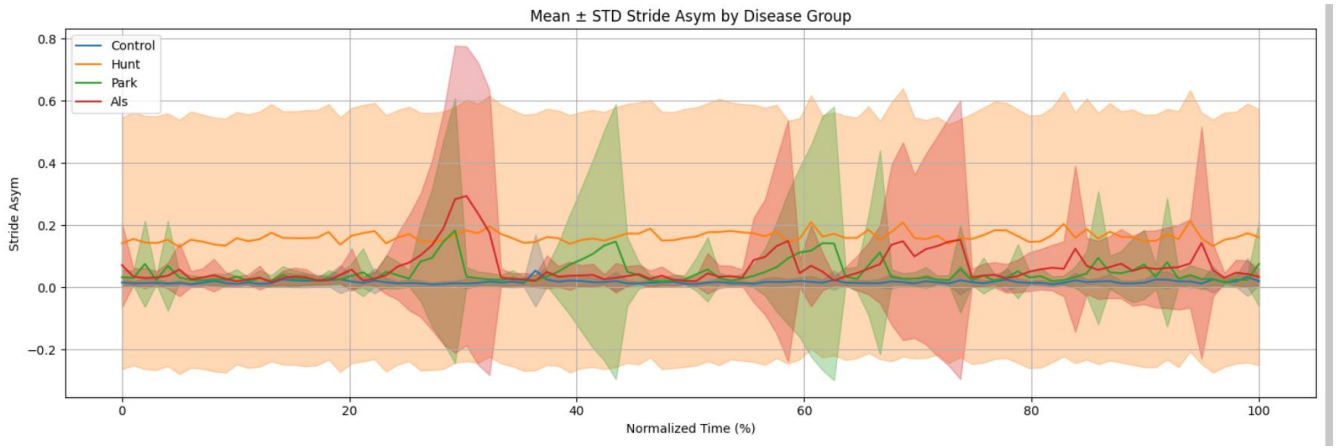
As shown, the stride intervals of people with ALS is higher compared to people with other diseases. People with ALS walk slowly with most of their time having foot on ground and hence, the graph shows longer stride intervals for people with ALS.



People with Huntingtons have more stride asymmetry (difference between left and right strides) As they walk unstably, they



As shown above, mean of percentage of stance in gait cycle is highest for ALS, followed by parkinsons, huntingons and then control. Which indicates, people with ALS or Parkinson put their foot more time on the ground.



The above shows asymmetric nature of the diseases. People with Huntingons have the most asymmetric nature of movement. ALS and Parkinsons have spikes at certain events indicating asymmetry at certain intervals (gait cycle for left and right legs at certain intervals are unequal)
Mostly, these diseases makes a person slow.

$$\frac{|XL - XR|}{0.5*(XL+XR)} \times 100 = Symmetry\ index$$

XL: value for left leg (e.g., stance time)

XR: value for right leg

Result difference = 0% then more symmetric

Elapsed Time	Left Stride	Right Stride	Left Swing	Right Swing	Left Swing %	Right Swing %	Left Stance	Right Stance	Left Stance %	GROUP	AGE(YRS)	HI
0	21.8733	1.0800	1.0700	0.3933	0.3867	36.42	36.14	0.6867	0.6833	control1	control	57
1	22.9500	1.0767	1.0733	0.3967	0.3933	36.84	36.65	0.6800	0.6800	control1	control	57
2	24.0133	1.0633	1.0833	0.3867	0.4000	36.36	36.92	0.6767	0.6833	control1	control	57

Columns - 'Elapsed Time', 'Left Stride', 'Right Stride', 'Left Swing', 'Right Swing', 'Left Swing %', 'Right Swing %', 'Left Stance', 'Right Stance', 'Left Stance %', 'Right Stance %', 'Double Support', 'Double Support %', 'Stride Asym', 'Swing Asym', 'Stance Asym', 'Unnamed: 0', 'GROUP', 'AGE(YRS)', 'HEIGHT(meters)', 'Weight(kg)', 'gender', 'GaitSpeed(m/sec)', 'Duration/Severity', 'SubjectID'

Method :

1. Combining the time series data with .tst data of the people containing age, height etc and converting into csv file.

2. Data preprocessing –

- Drop non predictive columns
- Encode any categorical variables into one-hot vectors.
- Convert class labels into numerical values.
- Scale X features with standard deviation 1 and mean 0

3. CNN model training –

1st layer -

- Conv1D with 64 kernels of size 3
- Each filter detects a different pattern in gait signal.
- Activation – Relu (non linearity)

2nd layer

- Reduces the size of the output from Conv1D by half.
- Focus on the most important features while reducing computation.

3rd layer

- Convolutional layer to extract high level features

4th layer

- A fully connected layer with 64 neurons.
- Learns non-linear combinations of features extracted by the convolutional layers.

5th layer

- Dropout layer – prevent overfitting

6th layer

- Softmax layer to classify disease

4. Early Stopping

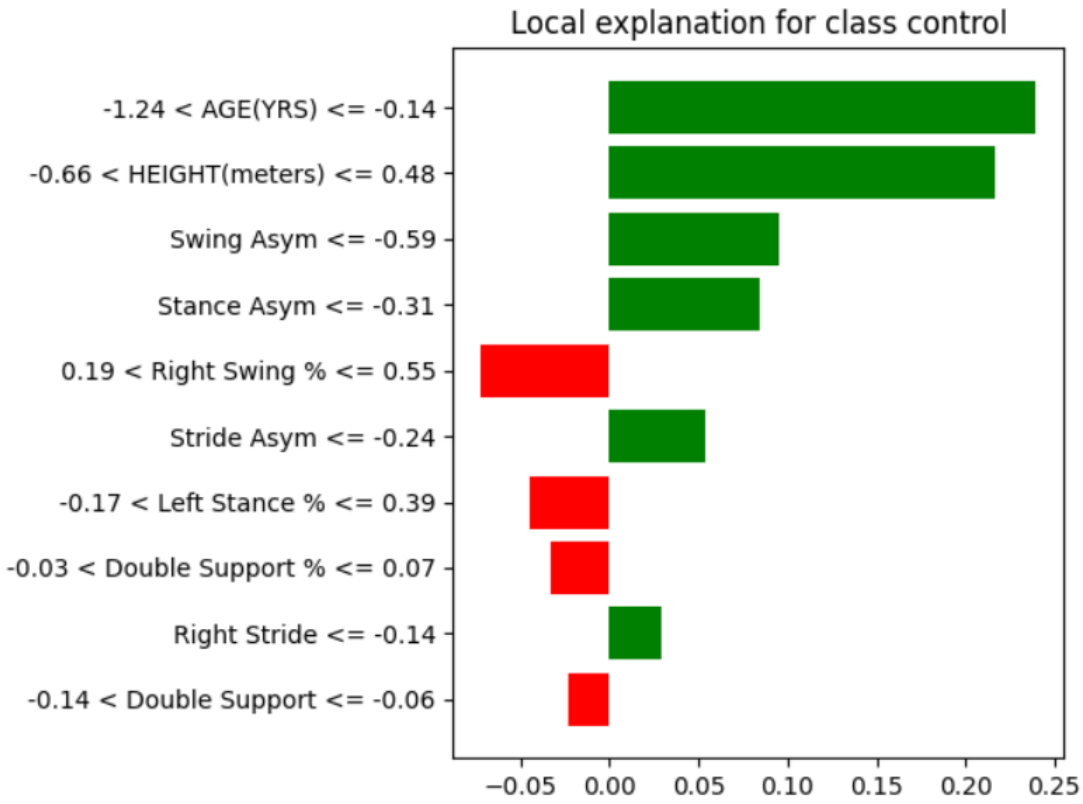
5. Compare with other models and use accuracy matrices for results.

Results :

```
266/266 ━━━━━━━━━━━━━━━━━━━━ 3s 8ms/step - accuracy: 0.997
Epoch 8/30
266/266 ━━━━━━━━━━━━━━━━━━━━ 2s 8ms/step - accuracy: 0.999
Epoch 9/30
266/266 ━━━━━━━━━━━━━━━━━━━━ 2s 6ms/step - accuracy: 0.999
Epoch 10/30
266/266 ━━━━━━━━━━━━━━━━━━━━ 3s 6ms/step - accuracy: 0.999
Train Accuracy: 0.9992
Test Accuracy: 0.9987
```

```
Test Accuracy Score: 0.9987

Classification Report:
              precision    recall  f1-score   support

         als       1.00      0.99      1.00       765
     control       1.00      1.00      1.00      1223
        hunt       1.00      1.00      1.00      1454
        park       1.00      1.00      1.00      1106

    accuracy                           1.00      4548
   macro avg       1.00      1.00      1.00      4548
weighted avg       1.00      1.00      1.00      4548
```

```
SVM Accuracy: 0.9990105540897097

Random Forest Accuracy: 1.0
```


Local explanation for class control
-1.24 < AGE(YRS) <= -0.14
-0.66 < HEIGHT(meters) <= 0.48
Swing Asym <= -0.59
Stance Asym <= -0.31
0.19 < Right Swing % <= 0.55
Stride Asym <= -0.24
-0.17 < Left Stance % <= 0.39
-0.03 < Double Support % <= 0.07
Right Stride <= -0.14
-0.14 < Double Support <= -0.06
−0.05
0.00
0.05
0.10
0.15
0.20
0.25