

CSE 572 Data Mining Assignment-1

Group Number: 34

Team Members:

Name	ASU ID	Mail ID
Manish Aakaram	1217852896	maakaram@asu.edu
Karishma Joseph	1217137207	kjoseph7@asu.edu
Gowtham Sekkilar	1215181396	gsekkila@asu.edu
Jerold Thomas	1215139965	jjthom25@asu.edu

Data Set:

The data considered for this assignment is a Continuous Glucose Monitoring (CGM) data series. The data is obtained from 5 subjects. There are also insulin basal infusion time series data and the insulin bolus infusion time series data. Each file is an array of time series data. Each patient has multiple such time series but the total number of time series data for each patient may vary.

Data Preprocessing:

The dataset contains many null values. Preprocessing had to be done to handle such missing values. Linear Interpolation was done for filling the missing data. Linear Interpolation makes a mean between the values before the missing data and after. We also used moving mean with a length of 15 to fill the missing values at the end which could not be interpolated.

Feature Extraction:

The four feature extraction techniques that we used are:

- Area under the Curve
- Root Mean Square
- Distance Travelled
- Roundness Ratio

The explanation for each feature extraction method and the features that are extracted are discussed below.

Area Under the curve:

Area under the curve computes the total area between the CGM graph and a line that corresponds to the minimum glucose level for the day. The intuition is that if the area is large then it correlates with increased blood glucose level.

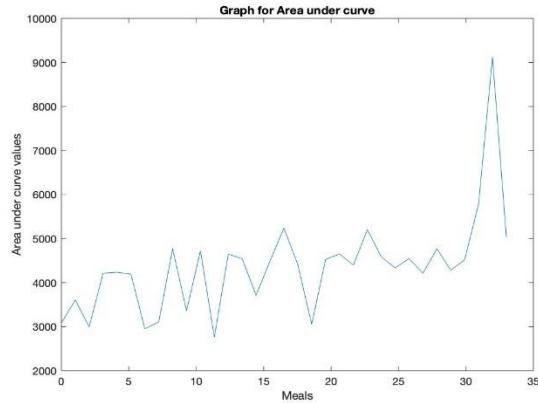


Figure 1 : Area under the Curve Graph for Patient 1

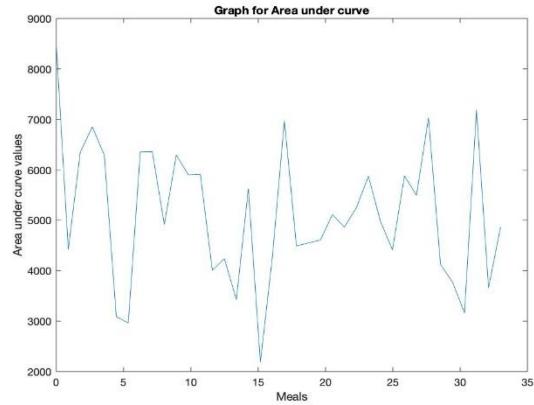


Figure 2 : Area under the Curve Graph for Patient 2

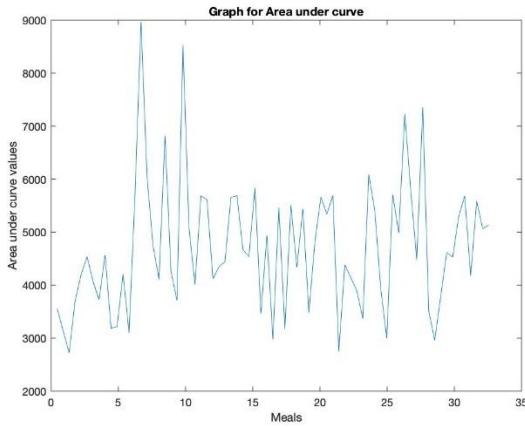


Figure 3 : Area under the Curve Graph for Patient 3

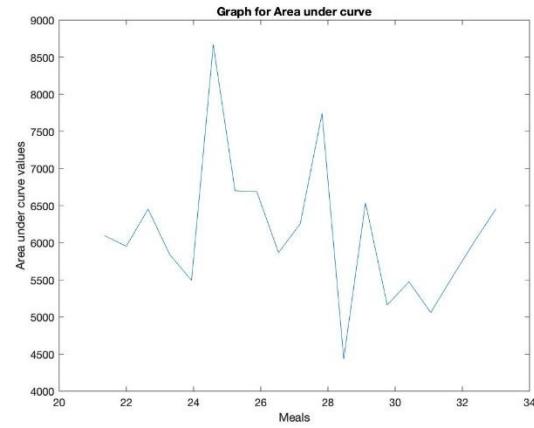


Figure 4 : Area under the Curve Graph for Patient 4

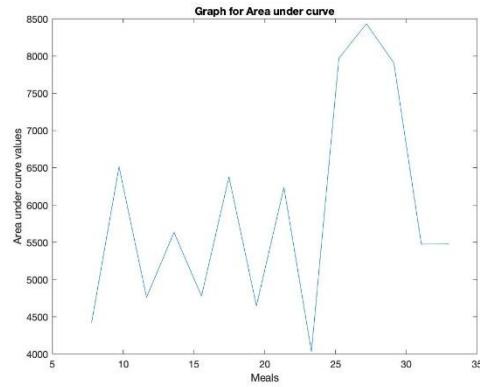


Figure 5: Area under the Curve Graph for Patient 5

From the above graphs, we can make an inference as follows

- 1) Days with low area values have acceptable blood glucose levels.
- 2) Days with high area values have excessive blood glucose level

For example, In Figure 1- On days 12 and 17, the area values are less which corresponds to acceptable glucose level. Whereas on day 33, the area value is very high, hence the blood glucose level is excessive.

This shows that the intuition is correct.

Root Mean Square:

RMS finds the root mean square of a set of numbers (i.e.) square all the numbers in the set, then find the arithmetic mean of the squares, and finally take the square root of the result.

$$RMS = \sqrt{\frac{\sum_{i=1}^N a_i^2}{N}}$$

Root Mean Square is also known as the quadratic mean. The intuition for using RMS is that we think that it helps in finding if there is a high change in blood glucose level.

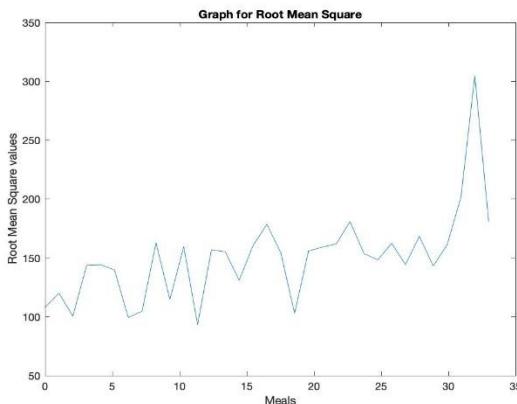


Figure 6: Root Mean Square Graph for Patient 1

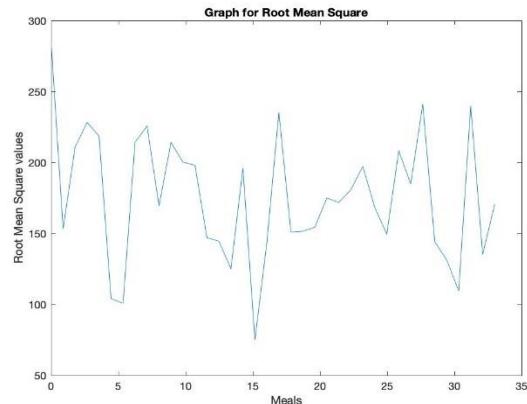


Figure 7: Root Mean Square Graph for Patient 2

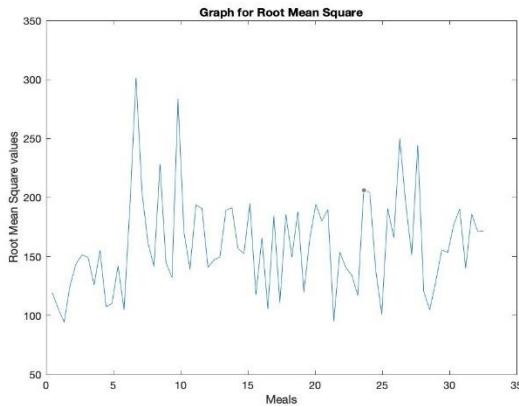


Figure 8: Root Mean Square Graph for Patient 3

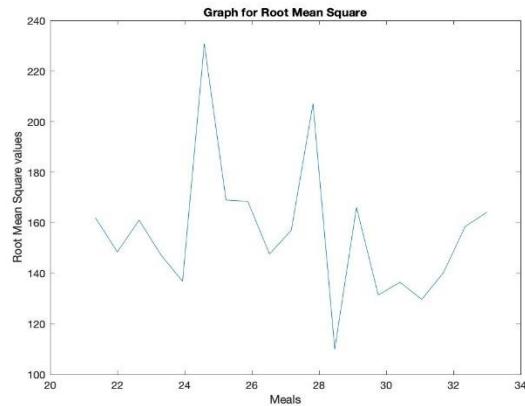


Figure 9: Root Mean Square Graph for Patient 4

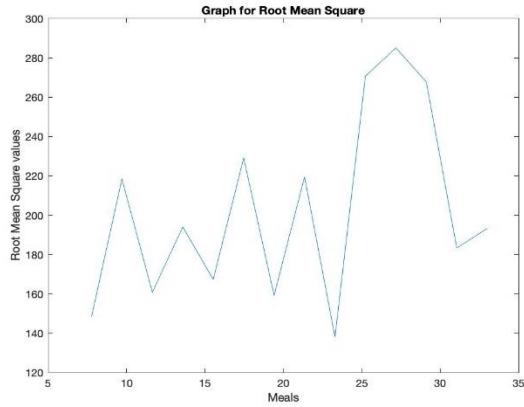


Figure 5 Root Mean Square Graph for Patient 5

From the above graphs, we can make an inference as follows

- 1) Days with low RMS values have allowable variations in blood glucose levels.
- 2) Days with high RMS values have too much variation in blood glucose level

For example, In Figure 7- On day 15 the RMS value is low which corresponds to acceptable variations in blood glucose level. Whereas on day 17 the RMS value is high, hence the blood glucose level has too much variation.

The result proves that our inference is correct.

Distance Traveled:

Distance traveled is a measure to gauge the total variability of the glucose levels in a day. This is found by taking the absolute difference of two consecutive glucose level values. The reason for using this feature is quite intuitive. The days which have allowable glycemic variability (refers to swings in blood glucose levels) will have a lower distance traveled values as compared to days with excessive glycemic variability.

$$DT = \sum_{i=0}^N |d_{i+1} - d_i| \text{ where } d_i \text{ represents the } i^{\text{th}} \text{ glucose level value}$$

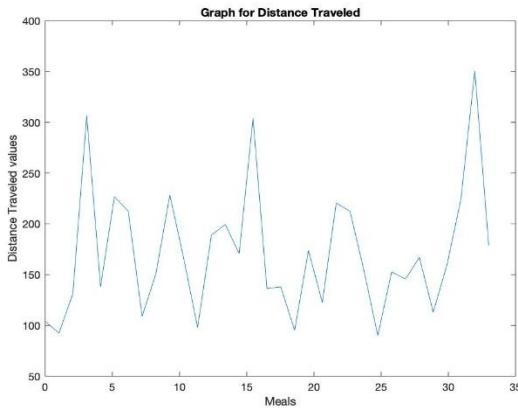


Figure 11: Distance Traveled Graph for Patient 1

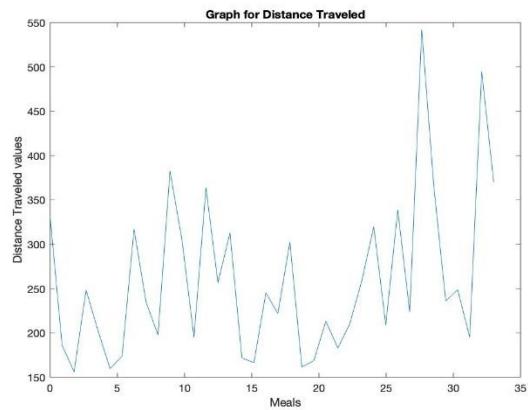


Figure 12: Distance Traveled Graph for Patient 2

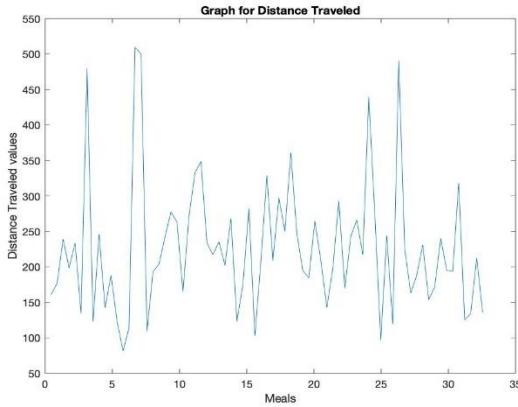


Figure 13: Distance Traveled Graph for Patient 3

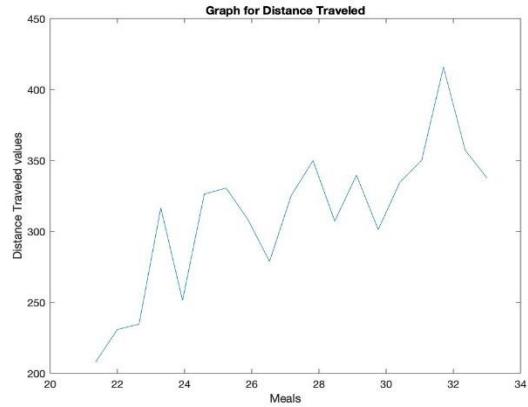


Figure 14: Distance Traveled Graph for Patient 4

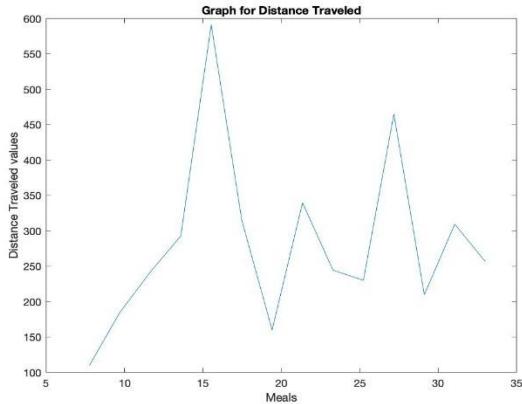


Figure 15: Distance Traveled Graph for Patient 5

From the above graphs, we can make an inference as follows

- 1) Days with less distance traveled values have acceptable change in blood glucose levels.
- 2) Days with more distance traveled values have excessive change in blood glucose levels.

For example, In Figure 13- On days 5 and 24, the distance traveled values are low which corresponds to acceptable change in blood glucose level. Whereas on days 7 and 26, the distance traveled values are high, hence the blood glucose level is excessive.

Roundness Ratio:

Roundness Ratio is defined as the ratio of the square of the perimeter of the CGM values and its area under the curve.

$$RR = \frac{P^2}{4\pi\mu_{00}} \text{ where } P = \sum_{i=1}^{n-1} \|p_{i+1} - p_i\|$$

$p_i = (x_i, y_i)$ is defined as a point in the graph.

For a perfect circle, the value of this feature is a perfect 1, and as the shape deviates from the circle, the value becomes larger. The intuition behind using this feature is that an acceptable day will have lower roundness ratio as compared to days where there is high blood glucose level. We'll use an example to illustrate this. Suppose the graph for an acceptable day resembles a rectangle. The graph for high blood glucose level may also be a rectangle (having same area) but will have more sharp points and spikes protruding from the edges. This will result in the perimeter being higher and hence the roundness ratio will be higher.

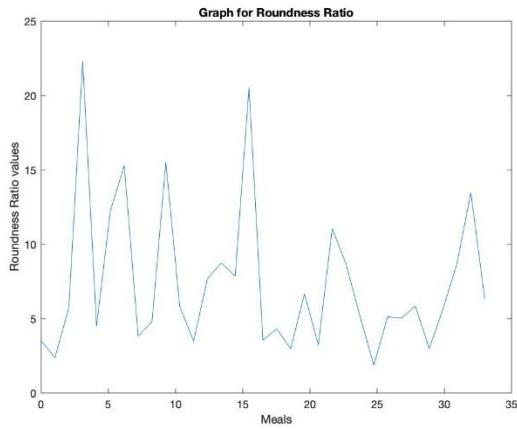


Figure 16: Roundness Ratio graph for Patient 1

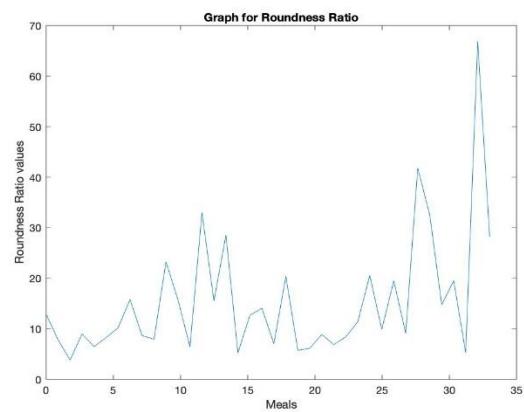


Figure 17: Roundness Ratio graph for Patient 2

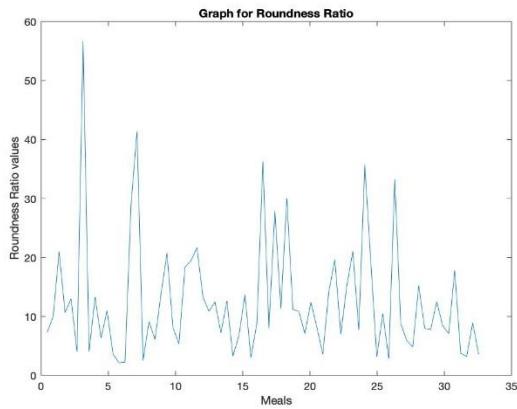


Figure 18: Roundness Ratio graph for Patient 3

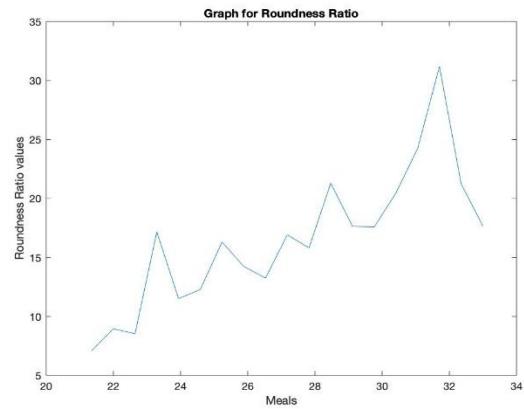


Figure 19: Roundness Ratio graph for Patient 4

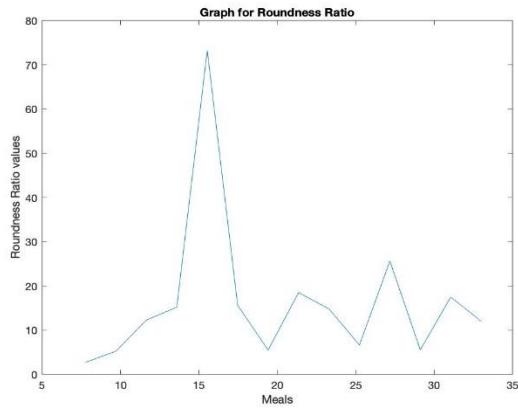


Figure 20: Roundness Ratio graph for Patient 5

From the above graphs, we can make an inference as follows

- 1) Days with low roundness ratio values have acceptable fluctuation in blood glucose levels.
- 2) Days with high roundness ratio values have excessive change in blood glucose level

For example, In Figure 20- On days 5 and 28, the Roundness ratio values are low which corresponds to acceptable fluctuation in blood glucose level. Whereas on day 15 the Roundness ratio values are high, hence the fluctuation in blood glucose level is more. This proves that the intuition is valid.

Feature Matrix:

We have considered multiple values for the same features. Distance Traveled and Rms features are taken for different window sizes. So in total we have chosen 10 features which consists of 4 distance traveled features (taken at 4 different intervals), 4 RMS features (taken at 4 different intervals), 1 Area under the curve feature and 1 Roundness ratio feature.

Feature Matrix Values:

The Feature matrix for patient 1 size is 33 x 10.

Feature values for Patient 1

Distance Traveled1	Distance Traveled2	Distance Traveled3	Distance Traveled4	RMS1	RMS2	RMS3	RMS4	Area Under Curve	Roundness Ratio
12	24	50	69	76.8205701619039	76.9403547029657	91.6867195697095	117.533747571418	3081	3.51054852320675
50.5714285714286		26	16	19	109.826179920447	123.026974130214	127.144155837523	128.581774475517	3605.71428571429
72.1428571428571		67	41	30	99.5765421929384	116.111623410015	112.56795271479	97.0018473961354	2997.42857142857
170.714285714286		96	89	50	170.904341306917	145.491955416473	113.741333336175	114.454040001772	4212.64285714286
50.2857142857143		60	57	42	107.923158604932	117.533360834662	143.183924948426	164.810083539702	4232.85714285714
144.857142857143		37	36	36	156.259817650748	140.8442079359	137.295633248443	136.65750154043	4192.57142857143
126.428571428571		35	40	45	106.818356215071	86.2101027616939	93.1909086474934	106.904375282518	2951.78571428571
23	46	54	44	83.6044257201735	93.0918412595374	115.630288576841	124.066697604738	3105.5	3.82579294799549
24.3333333333333		66	75	57	115.361364608973	129.074326719847	162.876975324658	194.452845417364	4765.333333333333
70.3333333333333		84	82	42	90.1371794051218	98.7218314254755	124.758311801805	139.429682505686	3360.166666666667
32	56	73	62	135.994485182304	129.637957404458	153.875687724628	185.048150981895	4723	5.8344272708246
61	31	30	15	95.5680868536839	87.0605498655998	92.218120877525	93.5433783672775	2761	3.47844983701557
24	74	79	71	121.27365748587	143.746936726255	169.938759022719	189.247312938772	4648.5	7.68441432720232
42.4	81	85	66	108.17752077026	133.179987576622	169.830396679639	193.742331226531	4542.7	8.7525326545887
49.8	75	88	52	83.4883464921902	89.2178131418731	128.018464577292	154.254217564265	3713.4	7.85640567242958
133.8	87	126	97	211.170272529066	207.958912025691	164.479620179965	110.967235131489	4488.1	20.5642565896482
37.4	63	91	47	131.397168919273	141.925396535709	176.480748370621	207.921751015748	5235.8	3.55341304098705
17	49	81	82	101.836142896321	107.957062171621	137.852358439284	178.185551908933	4411	4.31738834731354
56.1428571428571		30	13	30	89.3573198582237	99.2027309008266	101.599481029455	105.305184194235	2.96736643519448
30.7142857142857		85	92	48	104.30631890382	119.478792336472	160.444269565365	193.98960447714	4527.14285714286
20.5714285714286		62	88	48	111.689447366621	126.506198195259	159.71956631021	188.408405525675	4649.71428571429
43.5714285714286		72	134	126	82.5112360932295	87.9049072991533	135.359454114524	197.209855921876	4397.28571428571
80.2857142857143		120	118	38	111.559850022307	133.866283350894	187.465875682628	221.82302528441	5201.85714285714
52.1428571428571		65	61	59	130.50895660874	155.763515391997	178.575168791999	173.846588587859	4587.92857142857
17	59	62	32	105.4717971778203	120.701359486205	148.450481489773	171.403564193344	4334	1.86894323950162
12.66666666666667		66	113	89	93.5796155390929	102.691063610494	144.705343244942	194.204905845721	4545.333333333333
24.66666666666667		75	98	49	104.934476910329	114.372994269547	142.361894799518	173.978316308253	4214.166666666667
17	98	125	68	92.106460142598	120.325316613678	171.554391056916	209.408474257631	4768.5	5.84858970326098
52	47	40	21	118.699199660318	139.722126834917	157.746632293688	160.156457594101	4281.5	2.98236599322667
18	80	114	72	93.3182725943853	102.841901004855	149.388146055228	194.644197540958	4519	5.6469701261341
45.33333333333333		86	129	85	124.517780970341	146.62732990824	198.241358862464	249.668325438516	5787.833333333333
232.66666666666667		84	37	48	254.587073076733	318.163765721651	343.990618265303	337.44773736915	9122.666666666667
14	90	129	91	93.9696759598542	114.778838721176	166.44846322377	218.067128947195	5022.5	6.37949228471877

The Feature matrix for patient 2 has size 38 x 10.

Feature values for Patient 2

Distance Traveled1	Distance Traveled2	Distance Traveled3	Distance Traveled4	RMS1	RMS2	RMS3	RMS4	Area Under Curve	Roundness Ratio
148	100	115	133	323.517387477088	311.036829155181	284.876847395183	262.217467000198	8495.5	12.8963568948267
55	92	67	35	99.8208395075898	123.674425959305	155.770227053708	169.267136906027	4418	7.83069262109552
112.2	32	32	20	200.301083372008	216.231147365288	218.881661676308	212.842024216853	6347.9	3.84354510940626
159.4	39	54	62	233.627001864082	256.459459847857	249.968907157379	224.850657508085	6851.8	9.00530663475291
91.4	122	78	51	129.007968746121	185.49094635613	227.685388999023	256.041189868136	6293.2	6.50952774423187
101.8	58	26	36	118.243240821622	100.493780902104	88.303196688764	91.0459424486541	3086.1	8.2745342017433
37	64	55	46	74.291991493027	88.8440922474462	103.814957759205	110.833618135884	2964	10.2145748987854
177	105	100	62	241.985743381712	257.208793850538	223.403833286881	182.803421491751	6354.5	15.8138327169722
17	89	149	108	131.2265217098	156.168324107501	205.145624551749	262.771418121945	6359.5	8.68385879393034
82.727272727272727	99	75	53	108.173085973017	146.189789470587	177.608097685991	192.393063568027	4914.8636363636364	7.95466105934352
177.454545454545	134	110	147	202.453816792582	253.234171182614	257.778446098054	229.969760858011	6292.22727272727	23.2463757265786
92.8181818181818	117	112	129	173.820868180078	226.556635504037	245.366516201228	224.887850838024	5899.59090909091	15.7492486171824
96.09090909090901	96	84	58	173.592498028863	221.3690810855	228.182343353253	204.40178793027	5912.45454545455	6.43733706827023
151.63636363636364	102	88	121	74.5689707918694	107.095539335	131.368323288517	166.331815138514	4008.81818181818	32.9851340124148
84.636363636363637	69	63	44	161.018701752466	165.797795796292	147.588186272726	121.257427138974	4239.68181818182	15.5347089628393
111.909090909090901	65	130.5	135.5	115.206892585355	77.3815810830271	74.0294781581205	133.999915196717	3428.04545454545	28.5620772745957
42.8181818181818	66	84	79	125.937082546526	151.766866667866	184.345771951416	220.94199156422	5626.90909090909	5.24648383799218
109.54545454545455	62	44	24	84.284543645443	82.241330021081	74.689660961511	66.7628094621004	2184.27272727273	12.6986836778321
96.272727272727273	46	57	115	133.275034086448	154.700590584757	161.583921342328	160.695929687661	4271.63636363636	14.0832939938243
127	88	44	58	187.710948002507	238.360879950778	245.542997530705	244.429949065167	6974	7.06681961571552
156.6	115	102	78	164.420762408861	158.832959712111	134.851433403242	151.639224956653	4490.366666666667	20.3918224940799
75.2	42	45	42.5	145.187505584254	156.12582338381	155.612791598541	163.137072872645	4549.48333333333	5.74722184570409
67.8	67	56	55	126.821622762051	157.481600656885	177.042111066574	175.92637602672	4608.6	6.18266718743219
112.4	117	41	53	110.527086456113	168.938127458223	203.490227596494	208.540381962561	5112.3	8.90784187156466
48	67	76	70	96.6255659750565	129.344149813237	165.537965983089	196.861972873474	4863.5	6.88578184435078
100.5	93	51	30	116.813633622108	161.761890106753	188.40339094131	199.67769484392	5251.75	8.43723520731185
149	72	67	72	175.3530272399658	225.715830999149	225.672127412556	203.34989970917	5871.5	11.424848461211
135.5	110.5	56.5	44.5	134.781397084316	191.466232484525	200.076462656383	181.228760511028	4976.5	20.5767105395358
122	73	58	51	114.330223475685	162.4606246001	174.549029319454	159.237672564115	4412.5	9.89937677053824
191.66666666666667	239	111	46	102.521434073292	204.80877996455	244.162948570305	240.615498631776	5881.333333333333	19.5015491574095
88.66666666666667	77	73	83	157.947494792134	202.233887997401	222.62340806434	206.207089200065	5493.166666666667	9.10709265046472
141	116	105	200.5	208.225358686208	263.024713667746	297.005356547962	267.280963849586	7028.5	41.7961158141851
187.571428571429	121	121	87	102.012344191019	149.337019705582	183.959234614629	173.195370502899	4126.28571428571	32.5655132846658
134.142857142857	54	51	72	105.162061522554	108.523981430164	124.23438259267	135.308737875477	3772.07142857143	14.7832431160273
156.714285714286	63	50	59	121.813063539916	102.375867184694	109.65731746093	108.235433619988	3165.35714285714	19.5424380651505
102.285714285714	76	70	37	228.763214182843	252.555124495804	250.113610548923	239.406273176714	7186.64285714286	5.30658208041846
324.857142857143	139	90	111	117.709457251873	154.222389601982	177.097848450558	146.844382688861	3661.42857142857	66.881979227524
231.428571428571	81	84	95	142.306173227923	142.495294978147	158.834390597136	177.74784233443	4859.71428571429	28.2356777504346

The Feature matrix for patient 3 has size 75 x 10.

Feature values for Patient 3

Distance Traveled1	Distance Traveled2	Distance Traveled3	Distance Traveled4	RMS1	RMS2	RMS3	RMS4	Area Under Curve	Roundness Ratio
28.7142857142857		64	75	34	118.258226619823	126.88899451561	146.233002120215	156.991603712949	4138.35714285714
25		73	82	54	105.979904655033	118.149750894209	137.77748859533	136.887146617536	3560.5
45		87.5	94.5	57.5	111.641950000885	105.760127389553	115.246376397383	114.799231863125	3142.5
75		130	129	71	116.546986233021	95.1964476036981	93.2927942262716	93.1328474423888	2724.5
81.8		68	70	60	146.234004253457	136.466779174207	128.691668162882	111.881430744581	3701.6
87.6		51	97	105	165.83593096793	163.382109090307	162.395420893795	135.981616404571	4196.2
81.4		46	32	21	147.266535206709	144.278455520125	148.044526471538	153.562246780788	4537.8
328.2		147	116	54	221.452396990575	158.908492821841	102.320317938292	90.7964757025293	4063.9
51		76	45	29	95.3792430248846	120.166551086398	138.545693151787	139.09375856387	3726.5
76		100	92	116	152.771397846586	180.733606071578	184.919392758526	160.389299121629	4566
20.5		78	105	54	93.6174396164359	112.545344397068	126.58917081043	116.54147952688	3187.75
58		90	109	82	86.2015081074571	125.430531298477	138.954538020043	122.093481324017	3218
38.33333333333333		77	65	31	105.648210796229	135.5643693120748	154.881479613518	157.029528202177	4210.333333333333
18.66666666666667		45	52	29	124.139247531866	118.297102861628	100.078605469537	88.3335620351528	3099.166666666667
16		63	83	47	143.411296626172	163.340414194072	197.531585506539	224.95494498394	5722
373.6666666666667		133	76	78	356.793746831104	321.74283915179	291.81611775997	278.705612038614	8986.166666666667
420.833333333333		252	52.5	46	163.386181926271	193.656556133305	221.645960527554	226.668610731414	6050.333333333333
14		43	47	44	129.226253654117	135.798313016701	152.0819157268082	175.7994725450193	4744.5
79.5		63	82	55	113.529401478207	106.912878720777	137.949595537975	163.45586004123	4107.75
18		87	136	94	229.3232134463141	178.7818054091094	255.81296434558	230.595227094267	6811
97.25		65	85	74	146.756963207883	139.187577815629	159.163438078514	159.672107201545	4256.125
149.5		84	103	48	101.1717869196846	97.0524694079539	131.4644784374748	155.870575682636	3713.75
143.25		55	60	68	296.579426545403	307.27541687903	294.150548961829	283.24676039229	8535.375
68		51	50	50	178.920652804532	175.321263555068	173.103227216385	164.317043878871	5088.5
88.25		62	112	108	163.023483737773	160.01810785241	148.566483431485	116.881447318523	4009.625
172.5		92	107	105	224.798854534448	227.342758484586	205.718253929981	167.93413279090	5684.25
231.75		190	101	55	207.599870544276	227.13056917673	201.582376574577	170.714541107237	5613.875
100		73	99	68	125.061984636122	164.639515613411	164.630716232642	4119	23.935718441369
81.25		91	113	75	149.494864851619	147.60232777503	156.822242418504	147.447062919667	4345.825
100.5		82	88	64	148.238405954732	145.953292653382	162.538387074338	162.803283415627	4442.75
55.25		81	79	74	203.793170273196	204.499385748459	202.605976758293	184.166664209783	5652.875
120		142	86	82	196.921050169859	174.295464917792	165.548399953828	198.72288229612	5690
36.5		78	65	35	168.742481313983	171.211193346909	162.07736045032	146.668082695492	4670.75
45		66	85	61	155.71062905274	162.17807272479	160.961881427645	145.35536391266	4537.5
111.5		96	110	93	199.026664756201	190.352782533332	207.483624746593	212.646099508934	5837.75
22		31	47	54	137.947457361405	132.042692820439	119.910951805867	100.535837120176	3466.5
53.333333333333		92	100	89	159.907660118839	157.365238216211	175.493701606312	184.7762277501396	4938.166666666667
110.6666666666667		106	166	98	93.0004181591555	102.475717744609	131.807363237462	124.214989580017	2978.333333333333
72		74	97	73	155.096421622164	163.63122362731	188.3171698803094	206.780516050671	5462.5
120.5		88	134	97	108.443648961108	105.09216734069	127.781774201886	123.0868134138	3171.75
85		82	91	121	199.138896250833	202.005175451252	188.307050095297	161.958187645622	5506
141.857142857143		87	90	76	114.13266179825	132.087436305799	159.33555217292	176.08239708122	4338.92857142857
153.714285714286		67	54	51	136.07952655612	158.415736132953	165.230078060184	206.99077133141	5437.35714285714
87.5714285714286		50	74	63	136.697215359799	125.013817418134	130.715165705648	115.666448349417	3480.28571428571
79.5714285714286		66	45	64	209.813498330381	186.856775487148	160.246964864843	139.866494785693	4808.71428571429
79.3809523809524		72.66666666666667	135	151	212.452251004587	235.554014799546	227.472335989139	197.248240767328	5660.392857142857
60.1904761904762		70.333333333333	100	93	199.369065461617	197.195283407032	197.822004475097	197.29062079982	5340.07142857143
60		58	48	49	191.131387911132	180.114862341684	186.583639543918	200.25749782449	5690.5
123.83333333333333		37	48	37	100.971957052991	91.9421755037569	100.793688232328	98.5268768131085	2752.916666666667
131.33333333333333		126	116	108	202.170997432467	165.039451534591	146.243570172272	126.984968898377	4386.333333333333
79.5		68	74.5	48.5	158.815056506277	145.187104987135	149.49662527367	129.181250398599	4141.5
148.333333333333		56	53	49	120.373908210117	125.941183963713	135.278836616685	132.76500086448	3896.666666666667
141.16666666666667		63	94	60	125.64479920255	101.979488309892	106.63489100427	119.284910546565	3371.083333333333
39		86	122	105	223.669281718281	229.14543870735	231.55364585752	199.35806939309	6089.5
138.5		164	220	138	282.591262310358	275.823562050409	201.917624986023	108.472451467056	5407.25
115.5		132.5	114.5	77.5	169.994117545285	181.716124766076	148.683082732734	100.705466131313	3926.75
36		37	41	45	113.953476835644	101.314989363532	100.382450476348	95.5933622648085	2994.5
143		47	32	67	209.243637899614	202.318516287656	198.539344887839	191.829469723881	5702
48		26	22	40	174.095088960028	173.054116601505	174.249038082646	170.2877242482125	4986
236		111	206.5	184.5	300.86295599008	294.4677095623028	270.869349586576	221.660879395859	7231.25
59		72	120	111	202.576826710717	219.401458518397	22.7205436203649	198.36353205622	5787.5
73		78	61	44	168.338394388504	149.567558466519	152.614785451502	149.309927819467	4477.5
96.4		66	49	24	251.591923558766	263.735404594005	253.961879745044	240.429161746619	7351.8
108.2		77	91	65	129.79608622759	130.802766164801	134.80526030805	115.830988729943	3514.6
72.8		38	50	48	139.395064475038	117.225501412432	95.7235630182414	16.3748879865056	2959.9
77.9		31.5	36.5	40.5	151.170106171822	136.120852053006	124.161826222511	121.5528831760889	3788.533333333333
89		31	47	55	163.393084308976	155.177025009855	153.751274939282	168.189557551483	4617.166666666667
43.4		86	113	82	176.88401487998	170.078056946055	161.205684537713	147.12857704001	4530.3
66.8		42	63	79	190.187128902037	179.992175464937	167.605976500868	163.895671070657	5296.1
90.8		90	163	193	175.013610899267	192.883148797957	193.1796149609698	195.320984799114	5679.9
38.4		68	75	39	131.919733171349	145.07740804591	149.63895438437	142.037127028624	4175.7
46		47	26	36	174.062632405969	191.334785128058	198.073448268795	202.573892242268	5589
109.5		114	83	52	135.598764743636	182.297957900097	202.174542053507	187.06488908203	5064.25
45		67	64	40	150.740173809108	179.50993894996	189.014910426761	178.288326237943	5135
21.5		47.5	52.5	18	133.909017620174	154.783352992615	168.910479248624	167.386596617748	4640.60714285714

The Feature matrix for patient 4 has size 52 x 10.

Feature values for Patient 4

Distance Traveled1	Distance Traveled2	Distance Traveled3	Distance Traveled4	RMS1	RMS2	RMS3	RMS4	Area Under Curve	Roundness Ratio
64.66666666666667	49	67	62	110.776100511096	123.490522271586	143.702976498559	175.545901793131	4571.333333333333	5.78834281269749
23.5879120879121	30	46	72	131.068535269487	127.515417962621	140.185915516113	168.120086950857	4636.70604395604	3.24105001430109
48	81	103	105	176.691822108438	182.769850308573	205.626537021058	192.385503139353	5224.5	9.68992248062015
14	71	76	76	92.323886399994	113.236678927569	147.530274242888	182.655910986154	4478	4.89146940598481
48.5	151	163	104	117.779985566309	167.341512537033	244.045264057902	285.858833565228	6432.75	10.3878201391318
20	60	72	36	125.853883531657	137.260866435213	165.174563306934	185.463056845684	4746.5	3.50595175392394
97.5	78	113	79	188.548468569755	194.475518440567	230.71932918133	276.066526896557	7151.25	8.42793217968887
35	94	106	71	143.543373236106	174.787403332266	204.48960853794	195.741248498206	5230	5.26883365200765
59.1153846153846	72	97	53	175.118974350338	186.772588995281	225.10280479548	258.939445360564	6628.94230769231	4.6789708187637
35.7692307692308	98	120	61	123.108916514783	142.759174199705	192.654188544232	232.752892765931	5611.88461538462	6.01778769380494
139.346153846154	41	59	74	161.626886512693	140.451544539875	150.609066490338	177.12399354946	4857.32692307692	15.6084829451512
91.5384615384616	83	122	76	123.378241730641	141.902335172021	187.338051272414	235.724608195858	5621.26923076923	11.3454225261304
213.4230769230707	201	157	121	90.4125300592566	171.009834911223	238.730163383455	238.37632583204	5080.71153846154	44.1137836829802
58.3076923076923	50	67	74	93.3464280050489	86.1821538159939	109.789467949923	140.015908187086	3372.65384615385	11.310125744217
57.8076923076923	77	91	86	131.123150929494	133.595046179252	163.75591592367	165.22952330848	4312.59615384615	9.72643607980613
209.923076923077	110	59	40	238.424832162499	203.647422589319	168.891573609709	167.531001201677	5885.03846153846	15.082001645935
76.9615384615385	88	81	38	109.860891733794	138.688633756435	179.02258878161	199.973407322983	4727.48076923077	9.77443701484803
46.1538461538462	63	63	60	140.26127563867	153.766351680369	173.837698600003	176.772581172945	4688.42307692308	5.2677266867211
73.7307692307692	22	24	44	112.853579143259	114.801805497364	114.623494339273	125.889200923236	3438.36538461538	8.28011839945619
66.3846153846154	45	55	57	159.382339603592	154.454170902216	148.708989763106	141.119227476756	4259.30769230769	6.73610595938168
128.5	66	37	44	148.011232681848	124.993454374074	124.515789433235	136.841979204947	4188.75	10.578394509108
56.6153846153846	121	150	108	114.002257841235	114.611200468049	165.136857842766	223.494356243569	5134.19230769231	15.2282498644371
109.730769230769	55	33	51	115.633470743609	92.5526876973327	92.8273078943319	97.6896756431954	3028.13461538462	11.6384659839503
98.8461538461538	39.5	22.5	39.5	140.784069730708	128.002396994602	131.841157872232	137.440400306328	4021.28692307692	6.39283826604457
93.9615384615385	36	34	52	167.398488625444	163.617902998086	171.051294805229	177.46497733868	5015.51923076923	6.10336408502485
31.9230769230769	27	35	44	94.2497939978	94.249186155152	92.0316151212081	79.7393481059611	2464.46153846154	3.81181961804972
168.192307692308	63	49	40	183.501194311126	174.217471204031	183.89448555675	186.6479427838	5462.90384615385	11.275948347691
39.6923076923077	58	72	81	92.9244205200876	105.56815980037	137.881371937157	171.726791471485	4176.84615384615	8.07855081620713
120.576923076923	130	116	65	137.6688996200281	147.771322103998	177.333789017414	198.515170017993	5073.28846153846	14.1126234603269
37.5384615384616	60	94	68	107.472099035079	109.057949566441	145.158534024011	177.921075454565	4352.23076923077	5.55857772726222
39.3461538461539	56	100	111	78.792109996072	95.0865634327719	133.167358817938	165.49426258571	3981.67307692308	10.3849958259199
34.7692307692308	113.5	134	72	112.711960743634	122.24630519453	178.223748545065	221.75373683511	4969.11538461539	11.8606421558967
106.884615384615	38	22	17	148.024222889803	135.189398801898	126.502245758003	125.04147673816	4018.05769230769	5.0101298030622
30	93	123	79	91.587629246538	118.424889889883	169.611963320578	214.506198257554	4969.5	6.73890733474193
99.3333333333333	94	64	50	131.186931937259	160.303121956782	186.271795552042	180.886956160722	4813.833333333333	8.33710717954045
62.16666666666667	66	81.5	72	164.430642251106	183.339724406509	199.078673347544	193.346133327582	5302.666666666667	6.71268125052385
129	93	81	75	198.670078270483	191.405804034731	152.398103066219	127.020757072507	4675.5	13.9076034648701
81.93333333333333	36	44	77	165.074307846833	160.202428763793	160.29631652321	145.005955990529	4325.033333333333	8.51748451521247
89.86666666666667	161	137	75	144.771895675656	208.678961783188	280.21923235344	319.488369399918	7497.566666666667	9.64169945508247
72.7	183	150.5	122	147.405898796486	198.289159837574	204.638761546469	195.550272541127	5526.6	13.7037310462129
97.26666666666667	74	82	78	150.74179950867	188.976188972683	215.770500974784	218.127360628018	5510.633333333333	11.824605687146
68.33333333333333	65	84	84	114.424696246532	135.073380729951	162.29266493255	188.83237643085	4684.666666666667	10.2690574450927
156.4	69	72	35	137.948816595142	169.21398816056	188.994949247467	196.180807142057	5071.2	12.7621391386654
48.53333333333333	85	140	123	142.103090903908	154.620356180726	203.022390166924	264.034949477251	6552.233333333333	10.9236165446557
102.53333333333333	88	74	69	92.5761764410501	119.116062576113	136.828027431916	122.426972666825	3241.766666666667	14.0653196645176
48.4	102	103	73	123.31891978785	159.979260019428	207.60211069167	226.632060638143	5341.8	10.2855516866974
49.66666666666667	73	90	63	86.4379032086683	103.823276423318	139.153936997191	169.152432600142	3961.333333333333	8.60866150566588
136.7333333333333	50	34	23	116.774628142323	141.005802588533	149.881468318984	161.659293802513	4271.366666666667	10.0058539681549
117.8	93	56	42	78.8223572344801	112.088844631877	142.292592281474	155.347118649582	3852.4	11.534793894715
74.86666666666667	119	114	89	103.625102064016	156.045156634517	192.250310320118	180.492759864867	4331.433333333333	16.1965672128036
29.93333333333333	90	130	76	91.3685236342971	112.678545186981	160.806207478556	206.928402198354	4797.466666666667	7.92063959682425
43	73	64	45	127.33891785310	161.542282672089	185.821762107476	194.728482298255	5239	5.13380416109945

The Feature matrix for patient 5 has size 18 x 10.

Feature values for Patient 5

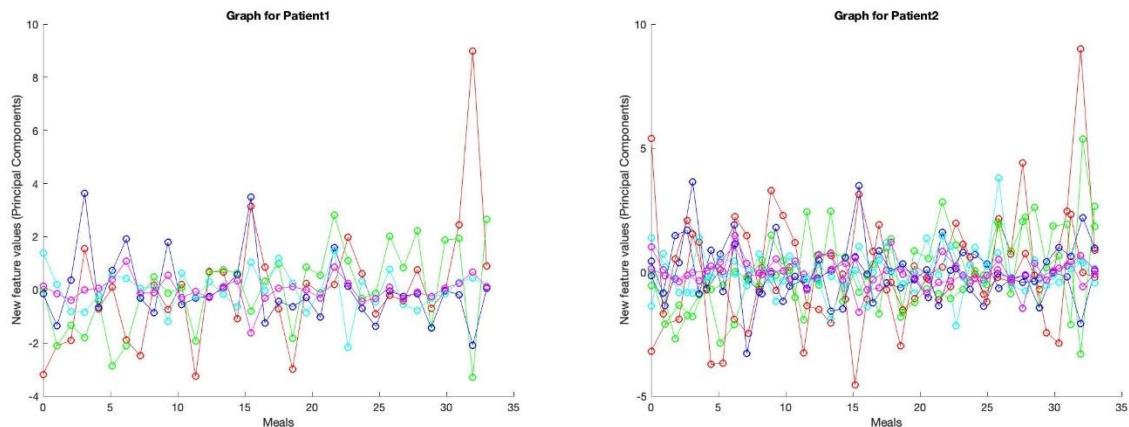
Distance Traveled1	Distance Traveled2	Distance Traveled3	Distance Traveled4	RMS1	RMS2	RMS3	RMS4	Area Under Curve	Roundness Ratio
92.8125	50	42	72	143.518956990444	154.501838765169	166.532825047141	184.438213354648	5302.40625	7.91115546008003
43.15	63	77	71	156.006641685539	170.387259446879	202.01822600437	234.280911417353	6176.075	4.25717344753747
143.75	124	146	142	94.1007770956223	130.792688147034	190.850488841158	257.869769809914	5822.375	22.8502222031388
57.7857142857143	97	73	18	151.050574379604	174.639628950591	210.881698848172	226.109185210074	5871.60714285714	3.77020945661194
57	38	43	36	159.524292820874	151.9584273292	145.396636199679	138.2625960606821	4417	2.73941589314014
71	59	70	57	238.993514556358	214.08643114059	205.582542591004	211.82475604312	6513.5	5.25447148230598
44	80	73	86	126.313894722631	143.756422781416	167.428410752992	178.66652532785	4760	12.3033613445378
78.75	140	169	108	123.286277622451	158.938810409999	215.230531798307	234.550246603618	5633.625	15.2126849941201
190.5	74	99	345	116.580551551277	159.967326209278	199.195655301287	217.446545155355	4776.75	73.2448317370597
136.75	149	128	78	108.995212058085	170.541383940566	233.303547882465	274.840449583522	6376.875	15.6343134372243
52	58	67	53	111.608691417828	131.283523857199	159.938908791621	179.066011190389	4646	5.5101622901421
54.75	101	148	148	137.218279576739	148.157471507975	197.346948752237	266.195075427436	6236.125	18.5099019824009
95.5	80	123	96	159.585791974098	161.304906085564	150.488598178793	115.463886522623	4031.75	14.8273702486513
99.25	59	77	68	204.015455909598	234.21939986416	263.236499265308	297.681495439795	7970.875	6.6510966237866
311	70	65	44	245.66771257127	241.24483037325	271.781161966756	297.605596232818	8434	25.6373013990989
110.75	85	69	38	196.861642403999	248.344518763753	280.422732835462	297.799963367725	7907.125	5.56397710925273
110.5	112	136	100	164.89974226784	169.205362046995	192.831768412497	197.446195202642	5473.25	17.5015301694606
141.75	123	78	58	108.122413263856	152.402278675401	193.740688925828	225.274781706092	5478.375	12.0328678668401

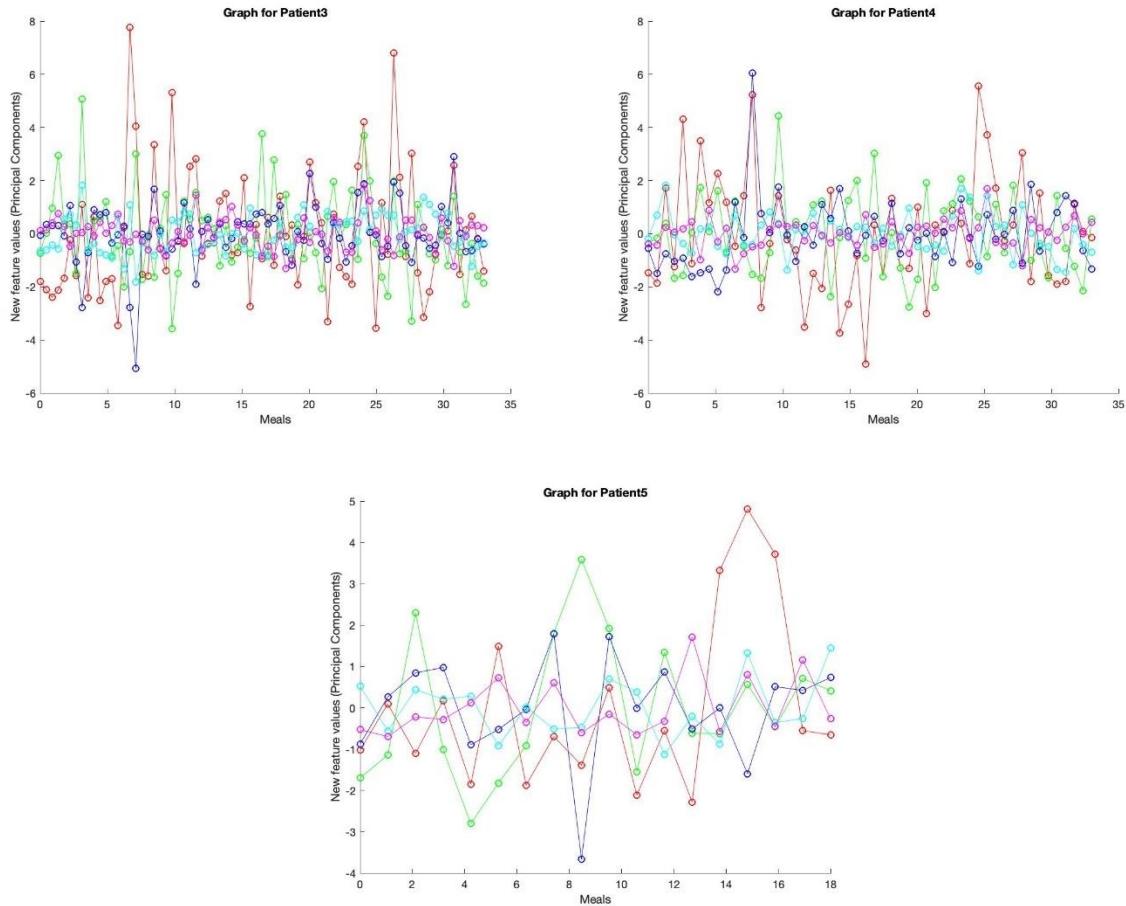
Principal Component Analysis:

Principal component analysis (PCA) is a technique used to emphasize variation and bring out strong patterns in a dataset. It's often used to make data easy to explore and visualize. It simplifies the complexity in high-dimensional data while retaining trends and patterns. PCA reduces data by geometrically projecting them onto lower dimensions called principal components.

New Feature Matrix Graphs:

The New Feature Matrix is calculated by multiplying the normalized feature matrix with eigen vector matrix. The below graph shows the plotting of the new top 5 features.





Results From PCA:

We have plotted the eigen vector graphs for each patient. From these graphs we find out the features which have high weights corresponding to the eigen vectors can be chosen as the top features.

On choosing top 5 best features:

The graphs below show the weights corresponding to respective eigen vectors. The X-axis represents the features in the feature matrix and the Y-axis denotes the weights. The top 5 features are the ones which have 5 largest weights. These 5 features can be retrieved by looking up at the features in the feature matrix.

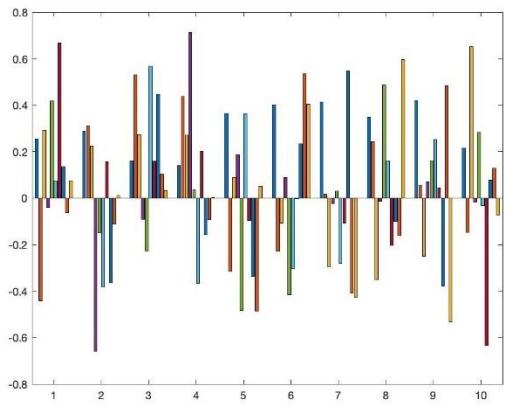


Figure 21: Eigen Vector - Patient 1

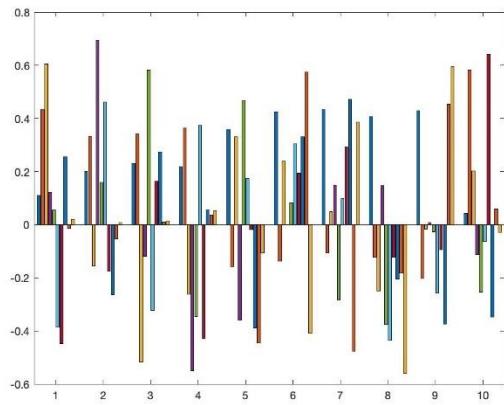


Figure 22: Eigen Vector - Patient 2

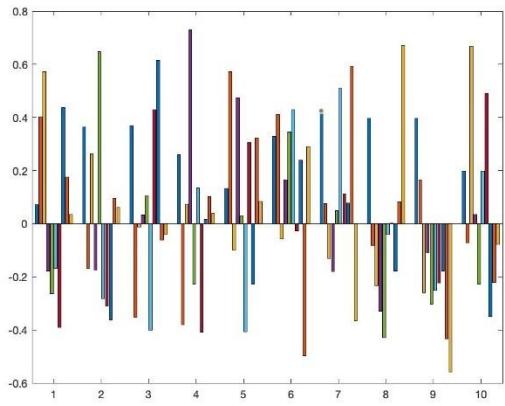


Figure 23: Eigen Vector - Patient 3

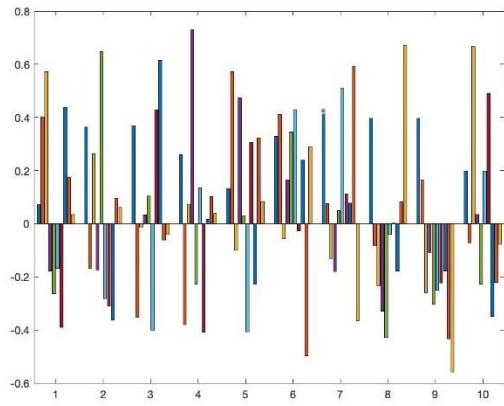


Figure 24: Eigen Vector - Patient 4

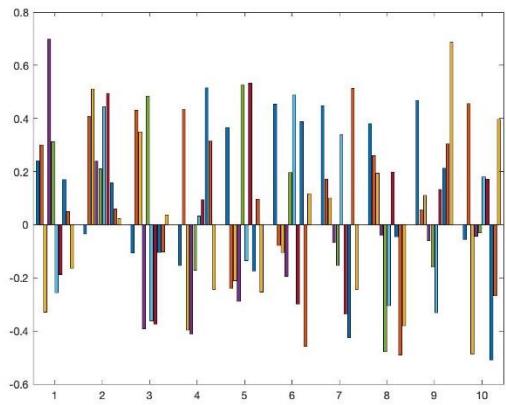
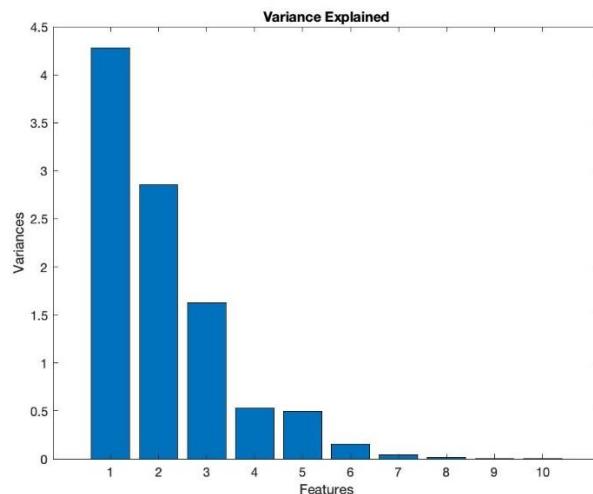


Figure 25: Eigen Vector-Patient 5

	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5
Top Feature 1	DT 4	DT 4	RMS 2	DT 4	DT 4
Top Feature 2	RMS 3	RMS 3	AUC	RR	RR
Top Feature 3	DT 3	DT 3	RMS 4	DT 3	RMS 3
Top Feature 4	RR	RMS 1	RMS 3	RMS 1	RMS 1
Top Feature 5	RMS 2	AUC	RMS 1	AUC	RMS 4

For example, let us consider, figure 21 - For Patient 1, the top 5 features are Distance Traveled 4 (DT 4), Root Mean Square 3 (RMS 3), Distance Traveled 3 (DT 3), Roundness Ratio (RR) and Root Mean Square 2 (RMS 2).

Justification for the top 5 chosen features: For each patient the top 5 features are chosen based on the order of variance. The PCA gives us the features by the order of their variances.



The above graph shows that the top 5 features are fair enough to represent most of the information. The first 5 principal components contribute ~97% of the information.

Conclusion:

PCA helped to reduce the dimension of the data and to determine which features are helpful. We hope the intuitions that we made for choosing the features are good.