ENGI 8103 – Engineering in Medicine

Implementation and testing of brain-computer interface algorithm

Abigayle Hickey
201732641

Table of Contents

Table of Figures

Table of Tables

Table of Equations

Introduction

A brain-computer interface (BCI) is a communication channel between signals of the brain and the outside world. For different tasks, the brain generates electroencephalography (EEG) signals, which can be collected through electrodes. EEG classification trains an algorithm to associate certain signal characteristics with different classes, to allow EEG signals to be sorted by class [1].

Motor imagery produces the same EEG signals as motor function, meaning a paralyzed participant thinking of moving their left hand will generate similar EEG signals to an abled participant moving their left hand. A BCI is able to classify the EEG signals of individuals, and given a certain class, such as left motor imagery, initiate a corresponding action. This action could then result in the movement of the left hand of a robot connected to this system [1].

This paper will be constructing a BCI to differentiate EEG signals between left, right and passive motor imagery classes. To construct a BCI, first the respective EEG signals are acquired and pre-processed. Signals are vulnerable to many factors, such as environmental noise, neighboring EEG signals, etc., and thus the acquired signals must be filtered to remove noise and ensure higher quality of data [2].

These signals are then analyzed further using various types of signal processing analysis, called feature extraction. Each trial, which are 1-2 second samples of data, are analyzed using each feature for each respective class. Optimal features show significant difference between differing classes, to allow the classifier to link certain feature ranges with different classes and make accurate predictions [2].

To ensure the features selected are most beneficial to the classifier, feature selection chooses the features that best differentiate the classes. Feature selection can be critical for the performance of the classifying algorithm. Redundant features that are similar between classes can decrease the classifiers accuracy, while an overload of features can lead to overfitting, which can result in the signal being rejected due to it not fitting all of the unnecessary information the algorithm has learned [2].

Features may also show the most results from certain feature-electrode pairs. Depending on the type of brain function taking place, different classes may have more activity in different parts of the brain, and thus different electrodes. Feature selection should choose the data points that have the most relevant data, without including redundant data [1].

With these features selected the classification algorithm is then trained. A training set of data, with the respective feature selection data points, is inputted into the classifier to learn. The accuracy of this classifier can then be tested using test data [2].

For BCI there are various types of classification algorithms, each with different benefits and disadvantages. In this paper we will be exploring the differences between using the Naïve Bayes, Discriminant Analysis and Error-Correcting Output Codes (ECOC) models. These will be used in conjunction with three different sets of feature selection inputs, analyzing the difference between using a standard, large and small set of data points on the accuracy of each classifier.

Methods

Data from 13 participants was extracted and pre-processed. Each participant conducted left, right, and passive motor imagery for 300 trials, in which each trial contained roughly 200 samples of data. Data was collected through 19 electrodes. Figure 1 shows an example of extracted data from participant 1, displaying the rows of samples and columns of 1-19 electrodes from which the data was collected. Column 20 indicates the trial number of the sample.

par1{1, 1}

	11	12	13	14	15	16	17	18	19	20
1	-11.5900	-12.2000	-16.9900	-22.6300	-11.3400	-24.5200	-4.1000	8.4700	3.7700	1
2	-8.0700	-10.9900	-15.2100	-18.4100	-14.3200	-26.2300	-5.4100	9.3000	2.5200	1
3	-7.9700	-17.6100	-13.6200	-10.0300	-14.3800	-25.9700	-7.0700	9.2400	2.6200	1
4	-7.9000	-18.6000	-13.1200	-5.1100	-11.1800	-23.8400	-7.9900	8.7900	3.3500	1
5	-8.8900	-14.8300	-10.4900	-5.7100	-7.3200	-19.9600	-7.8500	8.7400	4.7300	1
6	-9.6500	-11.0600	-10.6800	-4.0500	-3.5100	-17.0200	-7.7000	7.3700	4.7000	1
7	-9.1500	-6.5800	-9.9600	-6.3600	-2.0300	-16.8500	-6.9400	7.1900	4.0500	1
8	-12.9200	-2.8700	-9.7700	-6.2000	-0.8600	-16.3800	-5.3300	7.3800	3.5300	1
9	-14.2100	-1.1500	-9.4700	-0.8200	-0.7600	-15.4700	-4.5900	6	2.2100	1
10	-8.6000	-1.9000	-9.1600	-1.1800	-0.6200	-13.8500	-4.0700	4.7100	1.6000	1
11	-5.2300	1.8400	-11.1600	-0.7800	0.9200	-11.5000	-3.2500	4.2000	1.6600	1
12	-1.1600	4.2100	-9.1300	5.4500	0.9400	-9.5100	-2.6700	3.3600	2.5200	1
13	3.0100	1.5900	-5.4000	-2.8400	3.1300	-7.4100	-2.2900	2.1200	3.4800	1
14	4.8100	3.4000	-1.4100	-0.2900	6.3600	-5.5200	-1.9600	0.5200	4.0800	1
15	4.0200	7.6500	1.0500	1.7100	6.0200	-5.6300	-1.7900	-0.4400	3.7600	1
16	1.8100	8.9000	-0.1400	-1.4700	4.7100	-5.6600	-0.7800	0.0700	3.6500	1
17	2.5100	10.9900	-1.2000	11.9000	4.9500	-4.5200	0.6800	1.3500	4.3600	1
18	2.0600	11.1300	-1.3900	10.0800	6.4800	-3.1800	0.9600	2.2000	5.5500	1
19	-0.0300	8.4000	-1.5000	11.2000	7.3400	-0.8500	0.7200	3.6600	6.4600	1
20	2.2600	8.5700	-1.0800	12.4800	9.0700	-0.0800	1.4600	3.6500	6.6100	1
21	3.7900	12.0700	-2.4800	10.2100	10.8500	1.2200	2.2200	3.6800	7.4700	1
22	8.6500	15.6500	-2.9900	14.3200	12.5200	3.3500	3.1100	4.2300	8.5900	1

Figure 1: Processed data of participant 1. Rows indicate the data sample; columns represent the 1-19 electrodes. column 20 indicates the trial number

Feature extraction

For each feature analysis, the data from each trial for a specific electrode and class was analyzed. This was repeated for all 13 participants. Figure 2 shows the code used for extracting the data of a participant during left motor imagery. All the samples of a trial corresponding to one electrode are extracted and analyzed by the features.

```
% Left motor imagery data
columnNum = 1;
for trialNumber=1:300
    start = columnNum;
    % Collect all samples for each trial
    while ((leftData(columnNum, 20)) == trialNumber)
        columnNum = columnNum + 1;
        if (columnNum == length(leftData))
            break;
        end
    end

    for electrode=1:19
        % All samples (~200) used to extract feature information for
        % this trial
        trial(electrode, :) = featureExtractionFunc(leftData(start:(columnNum-1), electrode));
    end
    featExtraction{trialNumber, 1} = trial;
end
```

Figure 2: Feature extraction of the left motor imagery data. All samples from a trial are analyzed through each of the 27 features, for all 19 electrodes

The Jx-EEGT Feature Extraction Toolbox [3] was used to extract a total of 27 features, outlined in Table 1. All features of the toolbox were used except shannon entropy and log energy entropy, which both resulted in unusable infinite data for all classes. Features 20 to 27 used an additional opts variable which specified the sampling frequency of 200Hz.

Table 1: Names and descriptions of all 27 features used during feature extraction [3]

Representative feature number:	Feature name:	
1	Hjorth Activity	Variance of a time function
2	Hjorth Mobility	Mean frequency
3	Hjorth Complexity	Change in frequency
4	Skewness	Measure of symmetry of signal
5	Kurtosis	Measure of data distribution
6	First Difference	Difference between subsequent data points of the signal
7	Normalized First Difference	First difference applied to normalization formula

8	Second Difference	Difference between data points of the signal
9	Normalized Second Difference	Second difference applied to normalization formula
10	Mean Curve Length	Average length of signal curvature over time
11	Mean Teager Energy	Average frequency of signal amplitude
12	Log variance	Log root sum of variance of a signal
13	Arithmetic Mean	Average of the data points of the signal
14	Standard deviation	Amount of deviation between data points of the signal
15	Variance	Variance between data points of the signal
16	Median Value	Middle value of the data points
17	Maximum Value	Maximum peak of the signal
18	Minimum value	Minimum trough of the signal
19	Mean energy	Average energy of the signal
20	Band power alpha	Average power of alpha frequency range data points in the signal
21	Tsallis entropy	Probability equation of the signal
22	Renyi entropy	Quantifies diversity and uncertainty of signal
23	Ratio of band power	Ratio of alpha to beta band power in the signal
24	Band Power gamma	Average power of alpha frequency range data points in the signal
25	Band power theta	Average power of theta frequency range data points in the signal

26	Band power delta	Average power of delta frequency range data points in the signal
27	Band power beta	Average power of beta frequency range data points in the signal

Figure 3 shows the array of features returned for each trial. Collecting this data for each electrode produced a 19 x 27 feature extraction matrix, an example of which is shown in Figure 4. This displays the extracted feature data for each electrode and feature pair respectively, for each trial. Feature extraction matrices were created for every participant for all 300 trials of the 3 different motor imagery classes.

```
% Band Power Gamma
f24 = jfeeg('bpg', X, opts);
% Band Power Theta
f25 = jfeeg('bpt', X, opts);
% Band Power Delta
f26 = jfeeg('bpd', X, opts);
% Band Power Beta
f27 = jfeeg('bpb', X, opts);

% Feature vector
feat = [f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f12, f13
    f14, f15, f16, f17, f18, f19, f20, f21, f22, f23, f24, f25, f26, f27];
f = feat;
```

Figure 3: Feature extraction of data for 27 features, using Jx-EEG Feature Extraction Toolbox [3]

A large number of features were chosen for extraction to allow a large pool to choose from and analyze during feature selection. Though minimizing features during feature selection can be optimal, the more features to analyze through feature extraction will allow a higher chance of beneficial features to be selected.

featExtractionPar1{1, 1}

	1	2	3	4	5	6	7	8
1	31.8603	0.3732	3.3262	-0.3247	2.5309	1.5229	0.2698	2.4812
2	94.5123	0.2084	5.8470	-0.5143	2.4746	1.4159	0.1456	2.2928
3	34.4476	0.1753	5.8396	-0.6502	3.2975	0.8266	0.1408	1.4456
4	87.2050	0.1428	7.5075	-0.7665	2.7141	0.9132	0.0978	1.5877
5	4.6469	0.3716	3.0943	-0.2229	2.6121	0.5770	0.2677	0.9953
6	4.8039	0.3582	3.0509	0.1436	2.5290	0.5861	0.2674	1.0172
7	30.0342	0.1997	5.1469	0.1334	2.6146	0.8766	0.1600	1.4909
8	37.1228	0.2708	4.0091	-0.2978	2.0779	0.9910	0.1626	1.7694
9	74.6185	0.1817	5.5875	0.0893	2.2909	1.1337	0.1312	1.9606
10	86.4927	0.1844	5.1447	-0.2061	2.1574	1.1525	0.1239	2.0966
11	103.4192	0.2946	4.1401	-0.7011	3.3687	2.3294	0.2291	3.6931
12	231.4011	0.1517	7.9080	-1.0752	3.2787	1.7615	0.1158	2.9272
13	43.0495	0.3327	3.9213	-0.6877	3.3488	1.3827	0.2107	2.2110
14	142.1590	0.4773	2.7571	-0.7567	3.0207	4.3821	0.3675	6.5358
15	80.1445	0.1922	5.5291	-0.5559	2.8428	1.1840	0.1323	2.0859
16	327.1769	0.1244	9.4752	-0.6268	2.2172	1.1655	0.0644	2.1444
17	48.2299	0.1633	5.7215	-0.3377	2.3061	0.8782	0.1265	1.5569
18	16.3641	0.2686	4.0385	0.0141	2.1045	0.7202	0.1780	1.2637
19	17.2829	0.2498	3.6840	0.5015	2.4505	0.7773	0.1870	1.3879

Figure 4: Example of a feature extraction matrix, for participant 1, trial 1. The columns signify the 27 features, and the rows are electrodes 1-19.

5-fold cross-validation

Before selecting the optimal features, the data is randomized and split into five different groups, with one group designated for testing. To reduce the risk of the classifier specifically learning the inputted data, the testing data should not be used in either the training or feature selection process. This allows for the accuracy results to be a representation of the outcome of foreign inputted data, opposed to data seen and learned during training.

A 5-fold cross-validation procedure was conducted, with the list of training and testing groups outlined in Table 2. A total of 5 iterations were performed with the data. Each iteration randomized the 300 trials of feature selection data for each participant, within the three motor imagery classes. For each class, the 300 trials were then split into 5 groups of 60, and respective to the iteration, the training and testing groups were split up accordingly. This is done to ensure training of the classifier does not become dependent on the order of trials.

Table 2: 5-cross fold used for each iteration

Iteration	Training data consists of groups	Test data consists of group
1	2-5	1
2	1, 3-5	2
3	1, 2, 4, 5	3
4	1-3, 5	4
5	1-4	5

Feature selection

Feature selection was then analyzed using the training data of the first iterations 5-cross fold. A total of 240 trials for each of the 3 motor imagery classes were used, with 60 trials for each class reserved for testing. The Fisher score was used as the feature selection algorithm, shown in Equation 1. A higher Fisher score denotes a higher difference between the averages of the two sets of data for the two classes, as well as less variation within the set of data for each class. This is defined as the difference between the averages of the classes, divided by the sum of variation within each class [2].

Equation 1: Fisher score equation used for feature selection [2]

$$FS_f = \frac{\overline{c_{0f}} - \overline{c_{1f}}}{S_{0f}^2 + S_{1f}^2}$$

The Fisher score was calculated for the 3 types of class differentiations: left vs right, left vs passive and right vs passive motor imagery. An example of this is shown in Figure 5, depicting the difference between one trial of data through electrode 10, analyzed using feature 5, for both left and right motor imagery, shown on the left and right respectively. The larger the average difference between features of a given electrode, the more likely the classifier will be able to accurately predict the classes, using this feature-electrode pair. This is also dependent on how much this data point varies between both trials and participants.

The feature selection process took both the average and variance of each feature-electrode data point for 240 of the training trials for each participant. This created three feature selection matrices with respect to each class differentiation for all 13 participants. The 13 matrices for each class differentiation were then concatenated and averaged, to get the average feature selection between all of the participants.

featExtractionPar1{18, 1}

	1	2	3	4	5	6
1	72.1319	0.2766	4.0911	0.4506	3.3984	1.7686
2	33.1328	0.3416	3.2421	0.3163	3.1309	1.5018
3	29.9507	0.2536	4.8996	-0.0303	1.9806	0.7142
4	22.8793	0.3753	3.0909	-0.0214	2.6918	0.9335
5	5.6729	0.3763	2.6595	0.1747	2.1485	0.6606
6	4.9333	0.3979	2.6254	-0.1835	2.0620	0.6520
7	74.9878	0.1640	5.8568	-0.4882	1.9672	0.9956
8	75.6442	0.2136	5.3088	-0.1502	2.0010	1.0553
9	110.0343	0.1435	6.3707	-0.1207	2.1889	1.1633
10	55.5934	0.2505	4.2155	-0.4731	2.7653	1.1906
11	38.8917	0.4855	2.5862	0.0882	2.2489	1.8799
12	20.6249	0.6087	2.1652	0.1167	3.0378	1.1550
13	47.9653	0.2282	5.5771	-0.3564	2.0599	1.1262
14	40.2186	0.6776	1.9309	0.0156	2.3748	3.1293

featExtractionPar1{18, 2}

	1	2	3	4	5	6
1	370.3169	0.2086	6.2444	0.1831	2.6961	2.1449
2	81.6276	0.2267	4.7043	-0.1841	1.9669	1.5794
3	2.0669e+03	0.1872	7.4936	0.2360	1.5910	1.0156
4	3.1074e+03	0.1761	7.9626	0.2272	1.5941	1.1157
5	178.1355	0.1543	8.9252	-0.2718	1.7691	0.5607
6	181.5339	0.1548	8.9113	0.2840	1.7523	0.5895
7	95.2449	0.2291	5.8429	-0.0397	2.2715	0.8927
8	8.7214e+03	0.1780	7.9060	0.0800	1.7429	1.7552
9	713.2466	0.2135	6.4784	0.4176	1.7176	1.1377
10	816.5009	0.1809	7.3937	0.4951	1.9183	1.4511
11	8.8925e+03	0.1770	7.9522	0.2630	1.6654	2.5012
12	1.1520e+04	0.1864	7.5076	0.3226	1.6614	1.8698
13	1.0802e+03	0.1775	7.8957	0.3122	1.6550	1.3827
14	3.9574e+03	0.1786	7.6269	0.1400	1.6175	4.0406

Figure 5: Example of feature selection process. Data for each feature-electrode pair (seen highlighted in yellow for electrode 10, feature 5) is compared to calculate its Fisher score

These three averaged feature selection matrices represent the Fisher score for each feature-electrode pair (19 x 27 matrix) of the three class differentiations. For example, the left vs right feature selection matrix would depict the Fisher scores that are indicative of differentiating between a left and right motor imagery EEG signal.

To find the features that would be most beneficial for differentiating between all three types of classes, the three feature selection matrices were averaged, to get the average Fisher score for each feature-electrode pair, applicable to all three class differentiations. This is shown in Figure 6, where each feature selection matrix is concatenated to then find the mean of all three of the matrices.

```
% Combine feature selection of 13 participants into 3D matrix
LR = cat(3, featureSelectionPar1{1, 1}, featureSelectionPar2{1, 1}, featureSelectionPar3{1, 1}, featureSelectionPar4{1, 1},
LP = cat(3, featureSelectionPar1{1, 2}, featureSelectionPar2{1, 2}, featureSelectionPar3{1, 2}, featureSelectionPar4{1, 2},
RP = cat(3, featureSelectionPar1{1, 3}, featureSelectionPar2{1, 3}, featureSelectionPar3{1, 3}, featureSelectionPar4{1, 3},

% Calculate average of all 13 feature selection matrices of the three
% classes
averageLR = mean(LR, 3);
averageLP = mean(LP, 3);
averageRP = mean(RP, 3);

avgFeatureSelection = {averageLR, averageLP, averageRP};

% Caculate average between feature selection of the three classes
result = cat(3, avgFeatureSelection{1, 1}, avgFeatureSelection{1, 2}, avgFeatureSelection{1, 3});
result = mean(result, 3);
```

Figure 6: Feature selection matrices of each class differentiation type averaged together to create a final feature selection result

Figure 7 shows the resulting averaged feature selection matrix, depicting the Fisher score for each electrode (shown through rows 1 to 19) and feature (shown through columns 1 to 27).

19x27 double

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	-1.0436e-04	0.4735	-0.0321	-0.0345	0.0062	-0.0676	0.5438	-0.0542	0.3517	-0.0680	-0.0057	-0.3404	9.4320e-04	-0.0102
2	-8.6760e-05	0.4840	-0.0322	-0.0474	-0.0011	-0.0978	0.5641	-0.0791	0.3413	-0.0982	-0.0076	-0.5239	-0.0020	-0.0095
3	-0.0025	0.0595	0.0275	-0.0297	-0.0080	-0.0581	0.0038	-0.0693	-0.0794	-0.0584	-0.0136	-0.2784	0.0186	-0.0303
4	-0.0022	-0.1107	0.0073	-0.0904	0.0111	-0.2322	-0.2099	-0.1953	-0.1570	-0.2333	-0.0466	-0.8646	-0.0218	-0.0403
5	-0.0027	-0.6573	0.3073	0.0473	-0.0268	-0.4571	-0.7646	-0.3860	-0.9313	-0.4594	-0.1308	-1.1286	0.0709	-0.0213
6	-0.0019	-0.6661	0.3120	-0.0279	-0.0278	-0.4867	-0.8006	-0.3985	-0.9646	-0.4891	-0.1356	-1.1418	-0.0693	-0.0168
7	-0.0018	-0.1208	0.0101	0.0450	0.0016	-0.0443	-0.2564	-0.0403	-0.1226	-0.0445	-0.0105	-0.1845	0.0401	-0.0211
8	2.0071e-04	-0.8590	0.1258	0.1052	-0.0212	0.0028	-0.9818	-0.0329	-0.6749	0.0028	-0.0084	-0.0708	0.0409	-1.2303e-05
9	-4.6903e-04	0.1444	-0.0255	0.0934	-0.0037	-0.0214	0.0219	-0.0046	0.0232	-0.0215	-0.0011	-0.0641	0.0418	-0.0162
10	-6.5912e-04	-0.0841	0.0263	0.0823	-0.0143	0.0183	-0.1791	0.0217	-0.1208	0.0184	0.0043	0.0035	0.0408	-0.0176
11	0.0012	-0.1735	0.0501	0.0498	0.0101	0.0124	-0.3451	-0.0027	-0.3436	0.0124	8.9029e-04	0.1763	0.0284	0.0185
12	-1.2800e-04	-0.0433	0.0012	-0.0908	0.0042	-0.0349	-0.1195	-0.0307	-0.1080	-0.0351	-0.0034	-0.1959	-0.0073	-0.0075
13	-0.0011	0.1946	-0.0352	0.1120	0.0024	-0.0232	0.2232	-0.0247	0.1512	-0.0234	-0.0040	-0.1159	0.0401	-0.0232
14	0.0010	-0.2549	0.0666	-0.0095	0.0049	0.0099	-0.3726	-0.0056	-0.3274	0.0100	0.0039	-0.0897	0.0068	0.0109
15	-7.1976e-04	-0.0151	-0.0301	0.1359	-0.0062	0.0013	-0.0739	0.0014	0.0457	0.0013	-2.6864e-04	0.0049	0.0456	-0.0171
16	-0.0013	0.0118	0.0029	0.0873	-0.0148	0.0202	-0.1031	0.0016	-0.0739	0.0203	6.6130e-04	0.1022	0.0359	-0.0237
17	-0.0044	-0.2751	0.0689	-0.1025	-0.0038	-0.4388	-0.4674	-0.3574	-0.2790	-0.4410	-0.0792	-1.5911	-0.0108	-0.0602
18	-0.0118	-0.7353	0.1396	-0.0055	-0.0155	-0.6301	-1.1218	-0.5154	-0.8713	-0.6332	-0.1789	-1.5471	-0.0184	-0.0611
19	5.9516e-04	-0.2783	0.0908	0.0771	-0.0280	-0.0650	-0.4282	-0.0618	-0.3567	-0.0653	-0.0188	-0.2105	0.0404	-0.0011

15	16	17	18	19	20	21	22	23	24	25	26	27
-1.0436e-04	0.0063	-0.0017	0.0080	-6.4839e-05	-0.0013	0.2451	0.0141	-7.4221e-04	1.0966e-04	-6.9754e-04	-1.0147e-04	-0.0089
-8.6760e-05	0.0028	-0.0017	0.0077	-7.6958e-05	-0.0014	1.8338	0.0330	-3.7967e-04	0.0029	-1.4305e-04	-5.9371e-05	-0.0168
-0.0025	0.0212	0.0014	0.0141	-1.4344e-04	-0.0024	-0.3434	4.7136e-04	-0.0383	0.0262	-0.0076	-5.6922e-04	-0.0226
-0.0022	-0.0107	-0.0069	0.0116	-0.0019	-0.0132	7.9100	0.1119	-0.0261	-0.0805	-9.3973e-04	-0.0016	-0.0310
-0.0027	0.0657	0.0165	0.0279	9.3672e-04	-0.0378	3.9794	0.0354	-0.0584	-0.0841	0.0079	0.0207	-0.1473
-0.0019	-0.0652	-0.0268	-0.0155	0.0020	-0.0388	3.3333	0.0349	-0.0552	-0.1567	0.0089	0.0227	-0.1497
-0.0018	0.0354	0.0108	0.0183	-0.0011	0.0114	0.1389	0.0201	-0.0379	-0.0083	-0.0075	-4.7946e-04	0.0080
2.0071e-04	0.0371	0.0188	0.0197	4.7328e-04	0.0176	5.4620	0.0578	-0.0482	-0.0568	0.0265	0.0025	-0.0055
-4.6903e-04	0.0369	0.0074	0.0162	-5.3020e-04	0.0131	1.7227	0.0262	-0.0337	-0.0076	0.0065	-0.0011	0.0010
-6.5912e-04	0.0359	0.0091	0.0160	-4.8751e-04	0.0160	2.7234	0.0395	-0.0297	-0.0390	0.0113	-0.0011	0.0026
0.0012	0.0277	0.0134	0.0090	6.6797e-04	0.0099	-3.7042	-0.0413	-0.0268	0.0154	-3.5966e-04	7.6068e-04	-0.0044
-1.2800e-04	-0.0038	-0.0020	0.0032	-3.1078e-04	0.0028	3.9563	0.0703	-0.0276	0.0026	-0.0014	-5.2776e-04	-0.0083
-0.0011	0.0368	0.0101	0.0129	-4.1210e-04	0.0062	8.5226	0.0912	-0.0134	-0.0080	-0.0067	-0.0039	-0.0033
0.0010	0.0067	0.0050	0.0037	3.5059e-04	0.0018	0.3638	0.0116	-0.0132	-0.0081	0.0039	-7.4814e-04	-0.0045
-7.1976e-04	0.0385	0.0112	0.0169	-9.7048e-04	0.0076	6.5673	0.0745	-0.0297	-0.0300	-0.0013	-0.0013	0.0043
-0.0013	0.0302	0.0053	0.0144	-7.3283e-04	0.0050	4.8545	0.0733	-0.0168	0.0121	0.0045	-0.0022	9.3809e-04
-0.0044	-7.2047e-04	-0.0124	0.0131	-0.0032	-0.0100	1.1737	0.0368	-0.0801	-0.0963	-0.0039	-0.0024	-0.0541
-0.0118	-0.0179	-0.0182	-0.0017	-0.0012	-0.0563	3.6565	0.0472	-0.0647	-0.0330	-0.0473	-0.0026	-0.1770
5.9516e-04	0.0352	0.0148	0.0198	5.2488e-06	0.0122	2.5530	0.0277	-0.0188	-0.0354	-0.0010	0.0026	0.0014

Figure 7: Resulting feature selection matrix, showing the Fisher score for each feature (columns) with respect to each electrode (rows)

Using this matrix, the best feature-electrode pairs were found. The absolute values of the feature selection matrix were examined to analyze each Fisher score. Table 3 shows the three types of feature selection sets that will be used to train each classifier. As shown in Table 3, the data points were selected based on a minimum Fisher score value. These 3 sets of feature selection data points aim to analyze the effects of utilizing various amounts of data points on classifier accuracy.

Table 3: From feature selection matrix, resulting feature-electrode combinations to be used with out classifier algorithms

Minimum Fisher Score Value	Feature	Electrode(s)	Feature selection set	Total number of data
0.5 or higher	2	1, 2, 5, 6, 8, 18	1	40 data points, 7 features
	7	1, 2, 5, 6, 8		
	8	18		
	9	5, 6, 18		
	10	6, 18, 19		
	12	2, 4, 5, 6, 17, 18		
	21	2, 4, 5, 6, 8, 9, 10, 11, 12, 13, 15, 16, 17, 18, 19		
0.1 or higher	2	1, 2, 4, 5, 6, 7, 8, 9, 11, 13, 14, 17, 18, 19	2	97 data points, 11 features
	3	5, 6, 8, 18		
	4	8, 13, 15, 17		
	6	2, 4, 5, 6, 17, 18		
	7	1, 2, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 16, 18		
	8	4, 5, 6, 17, 18		
	9	1, 2, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 17, 18, 19		
	10	2, 4, 5, 6, 17, 18, 19		
	11	5, 6		
	12	1, 2, 3, 4, 5, 6, 17, 18, 19		
	21	1, 2, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19		
Higher than 1	7	8	3	21 data points, 4 features
	9	6		
	12	5, 6, 17, 18		
	21	2, 4, 5, 6, 8, 9, 10, 11, 12, 13, 15, 16, 17, 18, 19		

Classification

These feature selection sets were then used as the basis of the inputs to train each classifier algorithm. Depending on each feature selection set, the corresponding feature-electrode data points were extracted as the input data for each of the 240 trials of training data, for each class. This is shown in Figure 8, where Xiter1 is the input data for the 720 trials (240 for each class), which has a column of 40 data points with respect to the data points outlined in feature selection set 1.

Figure 8 demonstrates how the specified data points are extracted from the left motor imagery feature extraction training data. This was then repeated for the right and passive motor imagery. Xiter1 will then be used as the input to train each classifier along with the expected results of 240 'Left', 'Right' and 'Passive' classifications, respectively.

```
Xiter1 = zeros(720, 40);
% Feature selection iteration 1
for trial=1:240
    leftData = trainingGroup{trial, 1};
    Xiter1(trial, 1:6) = [leftData(1:2, 2); leftData(5:6, 2); leftData(8, 2); leftData(18, 2)];
    Xiter1(trial, 7:11) = [leftData(1:2, 7); leftData(5:6, 7); leftData(8, 7)];
    Xiter1(trial, 12:13) = leftData(18, 8);
    Xiter1(trial, 14:16) = [leftData(5:6, 9); leftData(18, 9)];
    Xiter1(trial, 17:19) = [leftData(6, 10); leftData(18:19, 10)];
    Xiter1(trial, 20:25) = [leftData(2, 12); leftData(4:6, 12); leftData(17:18, 12)];
    Xiter1(trial, 26:40) = [leftData(2, 21); leftData(4:6, 21); leftData(8:13, 21); leftData(15:19, 21)];
end
```

Figure 8: Example of feature selection process to be used to train classifier. Feature selection iteration 1 utilizes 40 data points, with 7 features in total.

Similar to Figure 8, the same is done for the testing data, except with 60 trials for each class, for a total of 180 inputs to be predicted. The training and testing data have the same data points for each trial, and thus the classifier will be able to learn from the training data the probability of each data points characteristics relative to the outcome. Extracting this information in the same order for both the training and testing data is how the classifier algorithm will be able to generate the probability of outcomes based on the input.

Table 4 shows the three types of classifier algorithms that were trained with this inputted data. The first classifier is the naïve bayes model. Bayes theorem generates a table of probability that, given a current outcome, predicts the likelihood of values for each data point. This probability table is based on both prior knowledge and the current input, thus as training proceeds, the probability of each feature with respect to each outcome continuously changes [4].

Table 4: Classifiers used in this project

Classifier name	MATLAB function	Classifier label
Naïve Bayes Model	fitcnb	1
Discriminant Analysis Model	fitcdiscr	2
ECOC model	fitcecoc	3

This model works with the assumption of independence among predictors and as such is most beneficial when there is no connection between various features. This makes it beneficial for multi-class prediction problems, since it is able to generate a link independently between one feature and outcome, without disrupting the probability of other outcomes [4].

The discriminant analysis model creates statistical properties of the inputted data for each type of class. Through training a probability table is created depicting the likelihood of certain data point inputs indicating an outcome. Unlike Bayes model, this type of analysis estimates the probability of each of the features as a whole and does not assume independence between them. The discriminant analysis model is beneficial for categorical output classification problems, for either binary or multi-class classification [5].

The ECOC model reframes multi-class classification problems as multiple binary classification problems. This uses binary classification models on each pair of classes to generate a table of probabilities between each pair. The outcome is then predicted through a tree of pair classifications, analysing each set of pairs individually to predict the most likely outcome [6].

These three classifiers were then trained with the input data, as shown in Figure 9. The classifiers were then used to predict the testing data, shown in Figure 9 as testIter1, similar to the previously shown XIter1. This returns 180 predicted outcomes to be compared with the expected results of 60 'Left', 'Right' and 'Passive' outcomes, respectively.

```
% Train classifiers with training data
classifier1 = fitcnb(Xiter1, Y, 'ClassNames', {'Left', 'Right', 'Passive'});
classifier2 = fitcdiscr(Xiter1, Y);
classifier3 = fitcecoc(Xiter1, Y);


labelsOne = predict(classifier1, testIter1);
labelsTwo = predict(classifier2, testIter1);
labelsThree = predict(classifier3, testIter1);

results(:, 1) = labelsOne;
results(:, 2) = labelsTwo;
results(:, 3) = labelsThree;
```

Figure 9: Each classifier is trained with the training data of the feature selection data points. The testing data with the same extracted data points is then used to predict the outcomes, which can be compared with the expected data to get accuracy results

This process was repeated for the other two sets of feature selections of different amounts of data points of the input. The results are then computed by finding the total amount of correct left, right and passive motor imagery, and then also the total accuracy. These values are found by comparing the results of each classifier to the expected outcome, and finding, for the left, right and passive outcomes, the correct value out of 60, and the total correct amount out of 180 outputs. These values are calculated for each feature selection set, each classifier and all 13 participants. For the 13 participants results, the mean and standard deviation of each outcome (left, right, passive, and total motor imagery accuracy) is calculated.

Results

Each iteration generated three matrices corresponding to each feature selection set, each displaying the percentage of accuracy for the left, right, passive, and total motor imagery predicted outcomes for the three classifiers, for all 13 participants. The average of these different matrices were then calculated for each of the 5 iterations. The average of each classifier with respect to feature selection set 1, 2 and 3 can be found below in Tables 5, 6 and 7 respectively.

Table 5: Results of classifiers for each participant for feature selection set 1, averaged over 5 iterations. Predicted percentage accurate for left, right and passive motor image classifications

	Classifier #1 (%)				Classifier #2 (%)				Classifier #3 (%)			
Participant	Left (%)	Right (%)	Passive (%)	Total (%)	Left (%)	Right (%)	Passive (%)	Total (%)	Left (%)	Right (%)	Passive (%)	Total (%)
1	99.33	0	66.67	55.33	71.33	19	67.33	52.56	99.33	0	66.67	55.33
2	100	0.33	66.67	55.67	66.67	22	68	52.22	100	0	66.67	55.56
3	99.33	0.33	66.67	55.44	66.67	20.33	68	51.67	100	0	66.67	55.56
4	100	0.33	66.67	55.67	68.33	26	67	53.78	100	0	66.67	55.56
5	99.33	1	66.67	55.67	69.33	31	69.67	56.67	100	0	66.67	55.56
6	99.67	0	66.67	55.44	67.33	23.33	69	53.22	100	0	66.67	55.56
7	99.33	0.67	66.67	55.56	63.33	22.33	68.33	51.33	100	0	66.67	55.56
8	100	0	66.67	55.56	61.67	23.67	68.33	51.22	100	0	66.67	55.56
9	100	0	66.67	55.56	63.67	26	68.33	52.67	100	0	66.67	55.56
10	99	0.67	66.67	55.44	65	24.33	67.33	52.22	100	0	66.67	55.56
11	99	0.33	66.67	55.33	63.67	26	67.67	52.44	100	0.33	66.67	55.67
12	99.67	0.33	66.67	55.56	67	20.33	69	52.11	100	0	66.67	55.56
13	99.33	0.33	66.67	55.44	67.33	27	67.67	54	100	0	66.67	55.56
Mean	99.54	0.33	66.67	55.5	66.25	23.95	68.13	52.78	99.95	0.03	66.67	55.55
± Std. Dev	0.37	0.30	1.48e-14	0.12	2.7	3.28	0.76	1.43	0.18	0.09	1.48e-14	0.071

Table 6: Results of classifiers for each participant for feature selection set 2, averaged over 5 iterations. Predicted percentage accurate for left, right and passive motor image classifications

	Classifier #1 (%)				Classifier #2 (%)				Classifier #3 (%)			
Participant	Left (%)	Right (%)	Passive (%)	Total (%)	Left (%)	Right (%)	Passive (%)	Total (%)	Left (%)	Right (%)	Passive (%)	Total (%)
1	99	0	66.67	55.22	60	24.33	69.33	51.22	96.33	3	66.67	55.33
2	99.33	1	66.67	55.67	61	26	67.33	51.44	96	3.67	66.67	55.44
3	98.67	0.33	66.67	55.22	58	23.33	71	50.78	99.67	1.333	66.67	55.89
4	99.67	1	66.67	55.78	60.33	28	69	52.44	96.33	3.33	66.67	55.44
5	99.33	1.333	66.67	55.78	65.67	30.33	70.33	55.44	89.33	100	68.67	56
6	98.67	0	66.67	55.11	61.33	27	720	53.44	89.67	400	690	54.22
7	990	1	66.67	55.56	56	26	68.33	50.11	90.33	9	66.67	55.33
8	100	0	66.67	55.56	55	28	69.33	50.78	96	5.333	66.67	56
9	99.67	0	66.67	55.44	58.67	29	70.33	52.67	97.33	2.67	66.67	55.56
10	99.33	0.67	66.67	55.56	56.67	25.33	70	50.67	96.67	3	66.67	55.44
11	99	0.33	66.67	55.33	54.67	26.67	70	50.44	82.33	13.33	66.67	54.11
12	99.67	0.33	66.67	55.56	56.33	24.67	70.67	50.56	93.33	4	67	54.78
13	99.33	0.33	66.67	55.44	54.67	30.33	70.33	51.78	87.67	8.33	670	54.33
Mean	99.28	0.49	66.67	55.48	58.33	26.85	69.85	51.67	93.15	5.46	67.05	55.22
± Std. Dev	0.4	0.46	1.48e-14	0.21	3.25	2.21	1.2	1.5	4.92	3.56	0.8	0.66

Table 7: Results of classifiers for each participant for feature selection set 3, averaged over 5 iterations. Predicted percentage accurate for left, right and passive motor image classifications

	Classifier #1 (%)				Classifier #2 (%)				Classifier #3 (%)			
Participant	Left (%)	Right (%)	Passive (%)	Total (%)	Left (%)	Right (%)	Passive (%)	Total (%)	Left (%)	Right (%)	Passive (%)	Total (%)
1	99.33	0	66.67	55.33	76.33	14.67	66.67	52.56	100	0	66.67	55.56
2	100	0.33	66.67	55.67	70.67	22.33	66.67	53.22	100	0	66.67	55.56
3	99.67	0.33	66.67	55.56	72.33	17.67	67.33	52.44	100	0	66.67	55.56
4	100	0.33	66.67	55.67	75.33	20	66.67	54	100	0	66.67	55.56
5	99.67	0.33	66.67	55.56	73.67	21.67	67.67	54.33	100	0	66.67	55.56
6	100	0	66.67	55.56	73	21.67	68	54.22	100	0	66.67	55.56
7	99.67	0	66.67	55.44	72.67	17.33	67	52.33	100	0	66.67	55.56
8	100	0	66.67	55.56	71.67	18.33	67	52.33	100	0	66.67	55.56
9	100	0	66.67	55.56	660	24	67.33	52.44	100	0	66.67	55.56
10	99.67	0.67	66.67	55.67	74	22.33	67	54.44	100	0	66.67	55.56
11	99.33	0.33	66.67	55.44	65.67	24.33	66.67	52.22	100	0	66.67	55.56

12	100	0	66.67	55.56	74.33	160	67.33	52.56	100	0	66.67	55.56
13	100	0.33	66.67	55.67	68.67	24.67	67.67	53.67	100	0	66.67	55.56
Mean	99.79	0.21	66.67	55.56	71.87	20.38	67.15	53.14	100	0	66.67	55.56
± Std. Dev	0.26	0.22	1.48e-14	0.1	3.3	3.3	0.44	0.87	0	0	1.48e-14	7.4e-15

Discussion

From the results in Table 5, the first set of feature selection showed comparable results between the first and third classifiers. The left motor imagery showed great results with an average of 99% accuracy, with passive motor imagery with an average of 66% accuracy. Both classifiers failed to predict the right motor imagery, with an average between 0.03 – 0.3 % accuracy. Though the results are similar, the third classifier showed less variance between participants. A classifier able to predict results with little variance is very beneficial, as to maintain accuracy for a large set of participants.

Improvements in the third classifier would be to increase its accuracy with the right motor imagery detection. To further identify where the classifier went wrong, a confusion chart can be used to visually show the falsely predicted outcomes. Figure 10 shows the confusion chart of an outcome of participant 1, iteration 1, using feature selection set 1 with classifier 3. The left, right and passive classifications are denoted as 1, 2 and 3 respectively. As shown in Figure 10, the right signals were most mistake for left signals, denoted as 40/60 right signals labeled as left. Thus, this classifier may see the most benefit in adding additional data points specific to left vs right signal differentiation.

The confusion chart can also be used to improve the classifier overall. As shown in Figure 10, all errors of the passive signal accuracy were mistaken for left signals, indicating a need for stronger left vs passive feature data points.

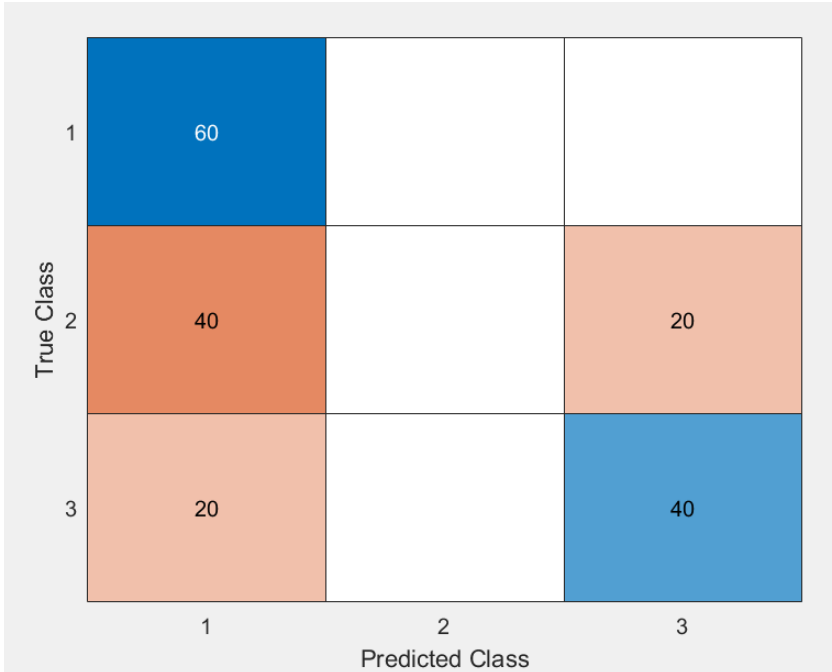

Figure 10: Confusion chart of results of first iteration using feature selection set 1, classifier 3. 1, 2 and 3 denote Left, Right and Passive, respectively.

Based on the apparent need for the classifier to have more data points for specific classification pairs, averaging the feature selection matrices of the three different class differentiations may have resulted in weak results. In future, it would be better to extract features from each independent feature selection matrix, to ensure a fair amount of strong features for each class differentiation. It's possible in averaging the three matrices, the differentiating features for the right signal were overpowered by higher Fisher scores of the other signals, and thus appropriate features for right differentiation were not included, resulting in an inability to predict the right motor imagery signals.

Classifier 2 showed the most variation in results, however also showed the highest amount of right signal detection. Based on the confusion chart in Figure 11, adding features that strongly differentiate both right-left and right-passive signals could raise the right motor imagery accuracy to acceptable levels for this classifier. Since this classifier shows the best results when considering the accuracy of all classes, this classifier would be recommended in future use with a different set of feature selection datapoints.

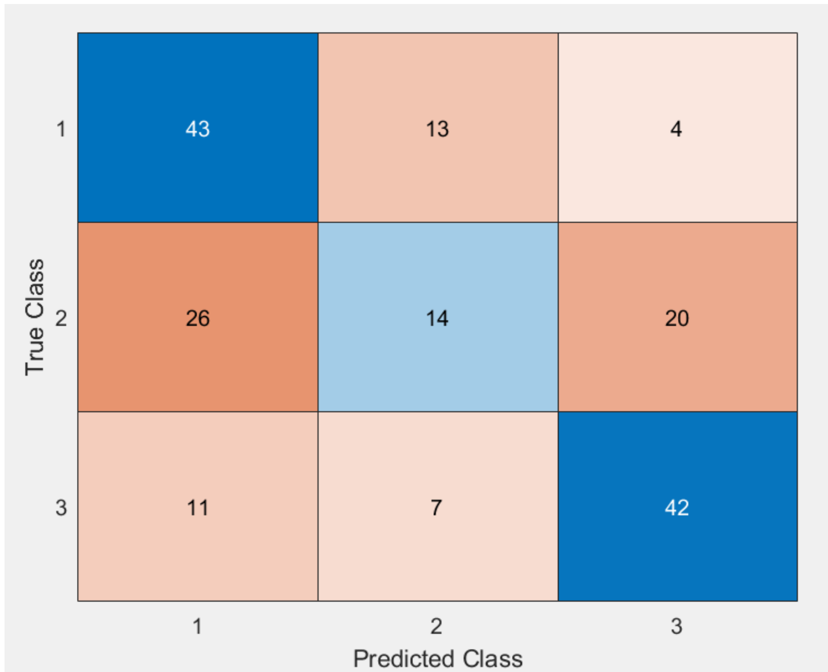

Figure 11: Confusion chart of results of first iteration using feature selection 1, classifier 2. 1, 2 and 3 denote Left, Right and Passive outcomes, respectively.

The results for the second set of feature selection, depicted in Table 6, used a much larger amount of data points compared to the inputs of feature selection set 1. This has various results for each of the different classifiers, but overall reduced the accuracy of the classifier and increased the overall variance.

The average accuracy of the left motor imagery detection decreased, while the right motor imagery detection increased. This could be confirmation that features that best differentiated the right signal were not properly used in the first set of feature selection, and their incorporation increased the accuracy of the right signals. However, the addition of these extra features also decreased the accuracy of the other signal classes, possibly due to an overload of data points and overfitting occurring.

The last set of feature selection data points only included a small amount of the highest-ranking feature-electrode combinations. These results are shown in Table 7. The reduced number of data points slightly decreased the average accuracy of the right signal and slightly increased the average accuracy of the left signal, with minimal change to the passive signal, and overall decreased variance.

This feature selection set was composed mainly of extracted data from feature 21, which showed comparatively large Fisher scores and was expected to perform highly. As shown in Figure 12, the Fisher score for this feature was comparatively much larger than all other scores, as very few scores were greater than 1.

18	19	20	21	22	23	24
0.0076	-7.5491e-05	-0.0020	0.5141	0.0083	-0.0014	-0.0011
0.0073	-6.8967e-05	-0.0018	1.3064	0.0187	0.0022	9.0281e-04
0.0153	-1.4429e-04	-9.2921e-04	0.2855	0.0037	-0.0195	0.0067
0.0118	-0.0018	-0.0067	4.1622	0.0553	-0.0142	-0.0287
0.0283	0.0014	-0.0193	1.1437	0.0097	-0.0244	-0.0079
-0.0161	0.0022	-0.0213	0.9385	0.0114	-0.0234	-0.0114
0.0188	-0.0011	0.0055	-0.3593	0.0049	-0.0155	-0.0017
0.0223	4.3150e-04	0.0059	2.3521	0.0261	-0.0209	-0.0258
0.0160	-4.9515e-04	0.0051	-0.1952	0.0036	-0.0159	-0.0122
0.0151	-4.8831e-04	0.0054	1.7157	0.0232	-0.0164	-0.0231
0.0095	7.1264e-04	0.0047	-0.9627	-0.0179	-0.0135	0.0083
0.0031	-2.1819e-04	9.7063e-04	3.1194	0.0442	-0.0119	-2.0782e-04
0.0130	-7.2249e-04	0.0018	6.0026	0.0602	-0.0038	-0.0045
0.0045	3.0176e-04	-8.6778e-04	-0.6887	-0.0067	-0.0059	2.4199e-04
0.0166	-8.3317e-04	0.0028	2.9328	0.0359	-0.0143	-0.0118
0.0144	-7.6745e-04	0.0011	2.6741	0.0407	-0.0088	0.0062
0.0128	-0.0034	-0.0053	1.1094	0.0224	-0.0388	-0.0680
-6.6674e-04	-0.0016	-0.0256	1.7689	0.0211	-0.0318	-0.0549
0.0216	2.8722e-04	0.0055	0.6032	0.0101	-0.0084	-0.0160

Figure 12: High Fisher scores highlighted of tsallis entropy (feature 21), from averaged feature selection matrix

Due to the low performance, this feature was further investigated, which showed very minimal differences between the feature extraction data between classes. An example of this is shown in Figure 13, comparing the results of feature 21 between data of left and right motor imagery, displayed on the left and right respectively. With a high Fisher score, the difference between the two classes is expected to be substantial, however Figure 13 shows very minimal difference.

19	20	21	22	23	24
174.2434	2.5831	0.9923	7.0285	1.2540	0.5228
147.7525	2.8402	0.9907	6.7446	0.7088	1.0063
121.5089	1.2419	0.9918	6.9318	0.6995	0.2796
253.8457	1.2999	0.9921	6.9845	1.4938	0.3429
7.9652	0.8720	0.9889	6.4878	0.7444	0.1258
6.4098	0.7253	0.9881	6.3890	0.9658	0.1501
30.7389	1.9255	0.9874	6.3154	0.6800	0.3831
451.1426	3.1409	0.9934	7.2488	0.6845	0.1539
139.4355	2.3650	0.9879	6.3686	1.0079	0.5670
142.9182	4.2780	0.9895	6.5711	0.7151	0.3635
356.4832	2.9329	0.9919	6.9412	2.1463	2.1967
1.0806e+03	2.9184	0.9931	7.1810	1.6911	1.4434
56.4557	2.5814	0.9892	6.5335	0.9116	0.6886
475.1814	0.4464	0.9921	6.9842	48.0616	9.4867
217.3067	3.6020	0.9912	6.8302	0.5323	1.0061
730.9705	1.2545	0.9918	6.9216	1.8952	0.4203
120.1092	0.7663	0.9906	6.7364	1.8783	0.2916
76.5424	0.8122	0.9915	6.8789	1.0578	0.3263
86.3866	2.3642	0.9909	6.7741	0.6823	0.1195

19	20	21	22	23	24
457.6327	8.7943	0.9931	7.1872	0.7170	2.4211
4.5325e+03	5.8930	0.9943	7.4586	1.0816	1.9041
5.2175e+03	0.3824	0.9949	7.6232	3.4884	0.3380
1.4331e+04	2.0757	0.9949	7.6154	0.6875	0.4518
84.1070	0.5291	0.9934	7.2342	1.8239	0.1373
84.6135	0.4509	0.9934	7.2538	1.6233	0.1146
91.7782	1.4318	0.9927	7.0907	1.3758	0.2350
6.7898e+03	1.1047	0.9949	7.6113	4.9568	0.2529
3.1662e+03	3.4717	0.9945	7.5186	1.5553	0.2984
1.7623e+04	4.5495	0.9949	7.6188	2.2036	0.2324
2.4877e+04	5.8517	0.9949	7.6056	1.0300	3.7551
6.1886e+04	7.8101	0.9949	7.6232	0.4889	1.6023
3.1436e+03	1.8224	0.9949	7.6016	1.1459	0.9960
1.7123e+04	5.9610	0.9948	7.5848	5.1317	5.2383
1.8590e+04	6.8816	0.9950	7.6371	0.2294	0.2173
6.2193e+04	9.5440	0.9950	7.6429	0.3651	0.3122
5.2333e+03	1.6905	0.9949	7.6108	0.8658	0.1676
2.8070e+03	1.0944	0.9948	7.5975	0.7918	0.1686
2.8272e+03	2.1183	0.9947	7.5512	1.5315	0.1820

Figure 13: Comparison of tsallis entropy (feature 21) results of participant 1, trial 1 of the left motor imagery (left), and right motor imagery (right)

Using the data from Figure 13, the two highlighted columns are compared with each other, calculating the difference of their averages divided by the sum of their variation to get the following Fisher score:

$$\frac{0.9908 - 0.9945}{5.3360x10^{-7} + 3.1623x10^{-6}}$$

Due to the minimal variation this feature generates for all classes, the division of these tiny values creates a large Fisher score, despite lacking strong differentiation abilities. Due to the lack of differentiation this feature provides, the inclusion of this feature could have skewed the results of the classifier and is not beneficial to use in future analysis.

This shows a flaw of the Fisher score and leaves other feature selection methods to be more desirable. Through manual inspection, feature 21, Tsallis entropy, is shown not to have beneficial differentiation between classes, and should not be used in future use for this classification problem.

Conclusion

A BCI algorithm was developed to differentiate between three classes of left, right and passive motor imagery EEG signals. 27 features were utilized in feature extraction and Fishers score was used as the feature selection algorithm to calculate the best differentiating features. Three sets of feature selections were created, denoted by feature-electrode pairs with a Fisher score of higher than 0.5, 0.1 and 1 respectively. Of these feature selection sets, the following 11 features were used in conjunction with certain electrodes as input data:

Hjorth mobility, Hjorth complexity, skewness, first difference, normalized first difference, second difference, normalized second difference, mean curve length, mean teager energy and log variance.

The feature selection sets were used to compare the difference in accuracy between classifiers with input data point amounts of 40, 97 and 21, to showcase the effects of the number of data points used in training a classifier. These data points were used to train three different classifiers, the Naïve Bayes, discriminant analysis and ECOC models. Classifiers were trained and tested with different iterative groups of data 5 times.

The results showcased a positive correlation between the variance of results and the number of data points used. Overall, the classifiers were able to adequately classify the left and passive signals, while failing to classify the right motor imagery signals. The feature selection set with the most data points showed the highest accuracy of the right motor imagery classification.

This is indicative of a flaw in the feature selection process. Features were chosen based on the average combination of the Fisher score of each class differentiation, which may have led to right signal specific features being overpowered by other features and not chosen. As such, the classifier received features helpful in differentiating only the left and passive signals accurately, without the right. In future, the best feature selections should be independently selected for each type of classification pair, to ensure that each signal has an equal amount of differentiating data points.

In addition to this, a flaw with the Fisher score was found. Features that have small variance both within and between different classes inadvertently create a large Fisher score, shown through feature 21, Tsallis entropy. As such, using an alternative feature selection algorithm would result in more accurate feature selection data points to utilize. Tsallis entropy, as well, was found to not be a suitable feature for this type of motor imagery classification, as very little variance is created between classes.

Three different classifiers were used, the Naïve Bayes, discriminant analysis and ECOC model. The discriminant analysis model showed the best ability to differentiate between all three classes, however also acquired the most variance between participants. The Naïve Bayes and ECOC models displayed similar results, however the ECOC model displayed the least variance of the three classifiers.

Though the ability to differentiate between classes is most desirable, minimizing variance between participants is also ideal for creating a classifier able to classify large sets of data, that comes from a variety of sources.

In future use, it would be recommended to utilize the discriminant analysis model, since this classifier had the most promise for differentiation of all three classes. Combining this classifier with re-analyzed feature selection data points for each type of classification pair could result in promising classification results.

This paper demonstrated the importance of feature selection when training classifiers. Feature selection is critical in both the amount of overall data points, but also importantly the amount of data points specific to differentiating each classification pair. Classifiers require equal data for each type of class to be able to best create a differentiating algorithm to accurately classify and differentiate each class.

References

[1] S. Park, J. Ha, D. Kim and L. Kim, "Improving Motor Imagery-Based Brain-Computer Interface Performance Based on Sensory Stimulation Training: An Approach Focused on Poorly Performing Users," NeuroHaptics, Seoul, 2021.

[2] S. Power, *Implementation and testing of brain-computer interface algorithm,* St. John's, 2022.

[3] JingweiToo, "Jx-EEGT : Electroencephalogram (EEG) Feature Extraction Toolbox," 10 January 2021. [Online]. Available: https://github.com/JingweiToo/EEG-Feature-Extraction-Toolbox. [Accessed 18 February 2022].

[4] Z. Zhang, "Naive Bayes classification in R," ATM, 2016.

[5] S. K. Dash, "A Brief Introduction to Linear Discriminant Analysis," 18 August 2021. [Online]. Available: https://www.analyticsvidhya.com/blog/2021/08/a-brief-introduction-to-linear-discriminant-analysis/. [Accessed 24 February 2022].

[6] A. Berger, "Error-Correcting Output Coding for Text Classification".