Feature Engineering for an Efficient Motor Related EcoG BCI System

Ritwik Jain
Department of Biotechnology
IIT Kharagpur
Kharagpur-721302 India

Prakhar Jaiman
Department of Electrical and Electronics Engineering
BITS Pilani - Goa Campus
Goa-403726 India

Veeky Baths
Cognitive Neuroscience Lab
Department of Biological Sciences
BITS Pilani - Goa Campus
Goa-403726 India

***Abstract*—Invasive Brain Computer Interface (BCI) systems through Electrocorticographic (ECoG) signals require efficient recognition of spatiotemporal patterns from a multi-electrodes sensor array. Such signals are excellent candidates for automated pattern recognition through machine learning algorithms. The importance of these patterns can be highlighted through feature extraction techniques. However, the signal variability due to non-stationarity is ignored while extracting features, and which features to use can be challenging to figure out by visual inspection. In this study, we introduce the signal split parameter to account for the variability of the signal and increase the accuracy of the machine learning classifier. We use genetic selection, which allows the selection of the optimal combination of features from a pool of 8 different feature sets. Genetic selection of features increases accuracy and reduces the BCI's prediction time. Along with Genetic selection, we also use a reduced signal length, which leads to a higher Information Transfer Rate. Thus this approach enables the design of a fast and accurate motor-related EcoG BCI system.**

***Index Terms*—BCI, ECoG, Machine learning, Feature extraction, Genetic Algorithm, Feature selection, Information Transfer Rate**

I. Introduction

Electrocorticographic (ECoG) signals are obtained via passing a low pass filter on the brain signals [1]. These brain signals are measured via electrodes placed on the brain's cortex region. ECoG signals are extensively used to find the focus of seizures in epilepsy (drug resistance) [2]. Brain-Computer interfaces (BCIs) use neurophysiological signals, in our case ECoG signals, to predict the action the user wants to perform [3]. ECoG signals have been growing its importance in the studies related to the brain-machine interface [4]. Hence, it is important to classify ECoG signals from the user's brain efficiently.

However, the appropriate feature extraction technique must be identified for accurate classification to capture the difference between the patterns from different data classes. To create a feature vector representing the data or feed to the machine learning classifier, features can be extracted from multiple feature sets and combined. For example, features based on Power Spectral Density features consist of one set of features.

Even though there is a significant amount of literature on EcoG data where different feature sets are discussed, the non-stationary nature of signals is often overlooked. Non-stationary signals are those whose statistical properties, such as mean and variance, change with time. If a signal is of length X, then it might have completely different statistical properties from length 0 to X/2 and length X/2 to X, so using a single value (feature) to represent this signal means a lot of information about the signal might be lost.

It is also difficult to manually identify the best feature set for a particular task by visual inspection or other methods. The best features would also depend upon the positioning of the electrodes. Using many different feature sets helps capture a wide spectrum of time-frequency relationships, but keeping a high dimensional feature set increases computational demand and may add unnecessary noise for the classifier. Hence, a mechanism by which the best combination of features for a particular task can be selected from a large pool of features is required. Feature selection solves two major problems in BCI research, which are the speed and accuracy of the BCI. An important metric by which both of these are accounted for is Information Transfer Rate (ITR) [5].

The key contributions of this paper are:

1) The introduction of a signal split parameter that considers the signal's non-stationary nature to increase the performance while using these individual feature sets.
2) We suggest Genetic selection as a method to select the optimal combination of complementary features from a pool of different feature sets.
3) The use of Genetic Feature Selection along with reduced trial length to increase the ITR of motor-related BCIs.

The eight feature sets used in the paper are Modified Stockwell Transform (MST), Relative Wavelet Energy (RWE), Autoregressive Coefficients (AR), Petrosian Fractal Dimension (PFD), Maximum Power Spectral Density (PSD), Frequency at max Power (FMP), Hjorth Complexity (HC), and Mean frequency (MFR). Out of these 8, 5 feature sets have been previously used for motor-related task EcoG data: MST [6], RWE [7], AR [8], PFD [9], HC [10]. But, all of these studies have neglected the effect of the non-stationarity in the signals. The signal split parameter tackles this issue by splitting the signal into different sub-segments and extracting features from each sub-segment.

To reduce the dimension of the Feature vector, channel selection has been previously made using different methods such as wrapper method [6], and cross-validation method [9]. Feature selection using the genetic algorithm in EcoG motor

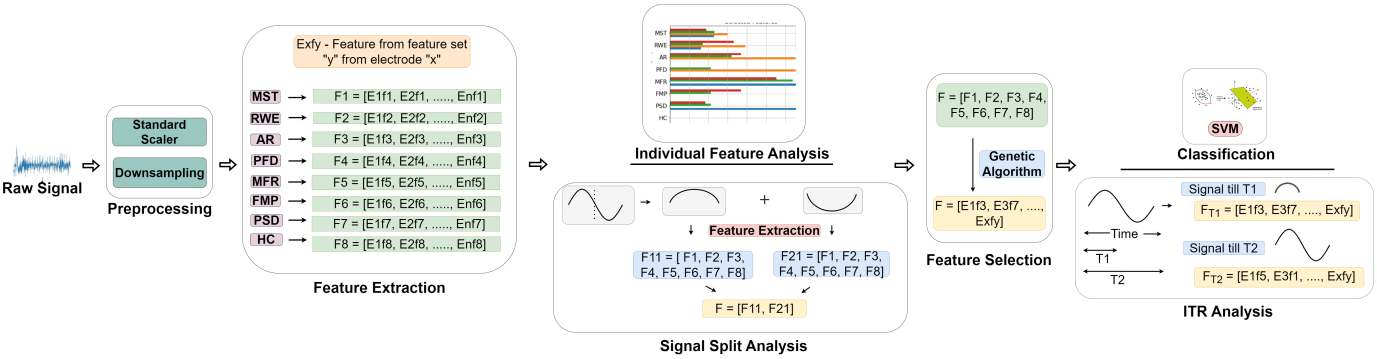

Fig. 1: The pipeline and main idea of the paper have been demonstrated in the figure. A standard scaler normalized the signal from the electrodes, then multiple features were extracted, and individual analysis, including the signal split, was done. Then these features were concatenated and passed through a genetic algorithm for feature selection. After feature selection, classification was done using either the whole length or reduced length of the signal

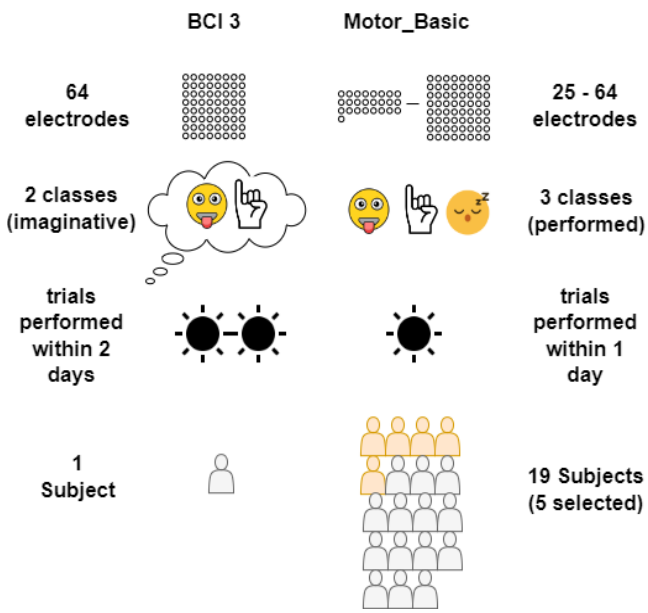

Fig. 2: Information about the datasets. BCI 3 dataset contained only one subject and 64 electrodes, whereas the motor basic dataset had 19 subjects, out of which five were chosen. They contain anywhere between 25 and 64 electrodes.

imagery has been made by Wei et al. [11] and Aswinseshadri et al. [12]. But, these studies create the final feature vector from only one type of feature set. Hence there are no different time-frequency relationships, and it does not allow the genetic algorithm the chance to select complementary features. This reduces the chance for the same feature pool to be generalized to other motor-related tasks.

The information about the Pipeline is demonstrated in Fig.1. We extracted various features from 8 different feature sets. For feature selection, we do a genetic search-based feature selection. The Genetic algorithm helps reduce the redundant data and only keeps the features determined to be most important. Different features can represent wide-ranging spatiotemporal patterns. Feature selection aims to find a pool of complimentary features accounting for most patterns. We also quantify the contribution of the features from each feature set in the genetic search. After Genetic Feature Selection, we get higher accuracy using a significantly reduced number of features, and using genetic feature selection on reduced trials increased ITR.

II. Methods

A. Datasets

The information about the datasets is given in Fig.2. The two datasets used were BCI Competition 3 Dataset 1 (BCI3) and Motor Basic Dataset. In the BCI3 Dataset, there was only 1 participant who had to imagine the action, while in the case of the Motor Basic Dataset, there were 19 subjects who had to act according to the instructions given to them. Out of those 19 subjects, we used five subjects in our analysis. All of these five subjects had a different number of electrodes and different positioning of these electrodes. The BCI3 dataset had a split

of test and training sets. Motor Basic Dataset had no test set and training set separation. These two datasets were publicly available, BCI3 dataset is available from BCI competition III [13], and Mot Basic is available from Stanford ECoG library [14]

B. Pre-processing

The sampling rate for both datasets was 1000Hz. The signal was downsampled to 100Hz from 1000Hz using the Fourier method. The average of each channel was taken, and the resulting signal was subtracted from each channel. The resulting data was then normalized by converting the data points on each channel into Z scores, which are subtracted from the mean and divided by the standard deviation [15].

C. Feature Extraction

Features can be used to represent a signal. As shown in Fig. 3, each feature set captures different signatures from the signal, so it isn't easy to pinpoint which signature would be appropriate for our task. If only a single feature set is used, then only one type of time-frequency relationship can be captured from the signal. Also, the features from that feature set might not be suitable for identifying the correct signatures from each electrode due to their signal's variation of spatial properties. For example, the features from the feature set Hjorth complexity might be good for electrode one but do not capture the pattern in electrode 24, whereas the features from the feature set Mean frequency do. Hence, we extracted a large pool of features from various different feature sets. The features were extracted from the signal of each electrode. Some of the feature sets: Petrosian Fractal Dimension (PFD) [16], Maximum Power Spectral Density (PSD) [17], Hjorth Complexity (HC) [18] have been previously used in EEG Classification, for PFD we used PyEEG [19] and for HC we used EEGExtract [20] to extract the features.

The features were chosen to capture various different time-frequency relationships from the signal as shown in Fig. 3, where these features are represented on a sine wave composed of various frequencies with a length of 300-time points. In the feature sets, MST, RWE, and AR multiple features of 35, 5, and 2, respectively, are extracted for a signal; hence error bars are plotted. According to the Fourier decomposition of signals, each signal can be represented as a linear sum of sine and cosine wave functions of different frequencies. Thus, we chose a sine wave to represent the features. The feature was calculated at every such 300-time point. Some features also vary with amplitude, represented in Fig. 4. The optimum signal split for each feature was also analyzed.

1) Modified S Transform (MST) - Stockwell-Transform (S-Transform) accurately captures the time-frequency information which is contained perceptual motion rhythms. The modified Stockwell-Transform feature was previously used for ECoG-based MI decoding using the BCI3 dataset [6], [21], [22]. [23]. The difference between standard S-Transform and MST is that MST has parameters to vary the height and width of the Gaussian window [24]. The mean value and standard deviation of the extracted features have a positive correlation with frequency and amplitude (Fig. 3, Fig.4).
2) Relative Wavelet Energy (RWE) - Relative wavelet energy captured the relative energy in different frequency bands and was previously used with BCI3 dataset [7]. We performed the 2nd-order Daubechies transform at the 4th Approximation coefficient (A4), and all four detail coefficients (D1-D4) were used as features. The mean value of the extracted features remains constant with change in frequency, but the variance decreases (Fig.3).
3) Autoregressive coefficients (AR Coefficients) - AR model is a simple temporal model where the present value depends on values at previous points. The regressive parameters of the AR model are AR coefficients [25]. AR modeling is a prominent feature extraction technique in signal processing with various applications such as speech signal processing, heart sound processing, and lung sound processing [26]. Hammon et al. [8] have used this feature, Haar wavelet decomposition coefficients, and spectral power estimates to classify the BCI3 dataset. The mean value and variance of the extracted features slightly decrease with increasing frequency (Fig.3).
4) Petrosian Fractal Dimension (PFD) - Fractal dimensions are used to characterize signals and measure their complexity [27]. The method proposed by Petrosian was a fast method to calculate the fractal dimension [16]. Fractal Intercepts, the intercept of the fitted straight line for fractal dimension, have been previously used to classify the BCI3 dataset [9]. PFD has also been used to classify epileptic signatures from EEG [28]. PFD features have an inverse correlation with frequency (Fig.3). In the case of changing amplitude, the feature value decreases and becomes constant (Fig.4).
5) Maximum Power Spectral Density The power spectral density is the power present in the signal in the frequency domain. PSD-based features have recently been used to classify EEG data [29], [30]. The PSD Features oscillates between different values with changing frequency (Fig.3). It has a direct correlation with amplitude (Fig.4).
6) Frequency at max Power - The frequency at maximum Power in the frequency domain. The feature value increases with frequency, as shown in Fig. 3.
7) Hjorth Complexity: Complexity is one of the parameters from Hjorth Parameters. The parameter of complexity helps in identifying the frequency change. This parameter measures the similarity between a sine wave and the signal. The closer the signal is to the sine wave, the closer the value is to 1 [18]. Hjorth complexity was previously used to classify EcoG data in non-human primates [31]. Hjorth Complexity and other Hjorth parameters are commonly used as a feature extraction technique for classifying EEG data [32], [33]. [34], [35]. The HC feature value shows a very complex relationship

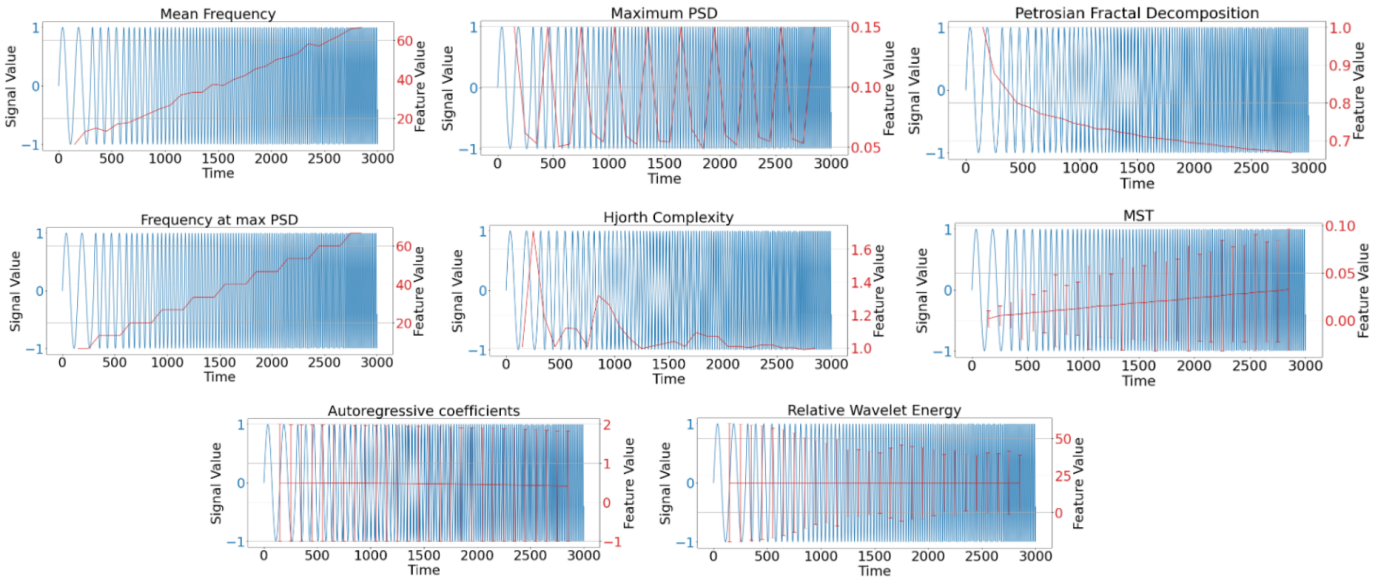

Fig. 3: Feature sets represented on a sine wave of different frequencies: Sine waves of frequencies 20, 40, 60, 80, 100, 120, 140, 160, 180, and 200 were concatenated. Each sine wave was of length 300, and the features were calculated. The red line depicts the feature values, and the blue line wave depicts the concatenated sine wave. The feature sets MST, AR, and RWE extract multiple features per signal; hence error bars are plotted for them.

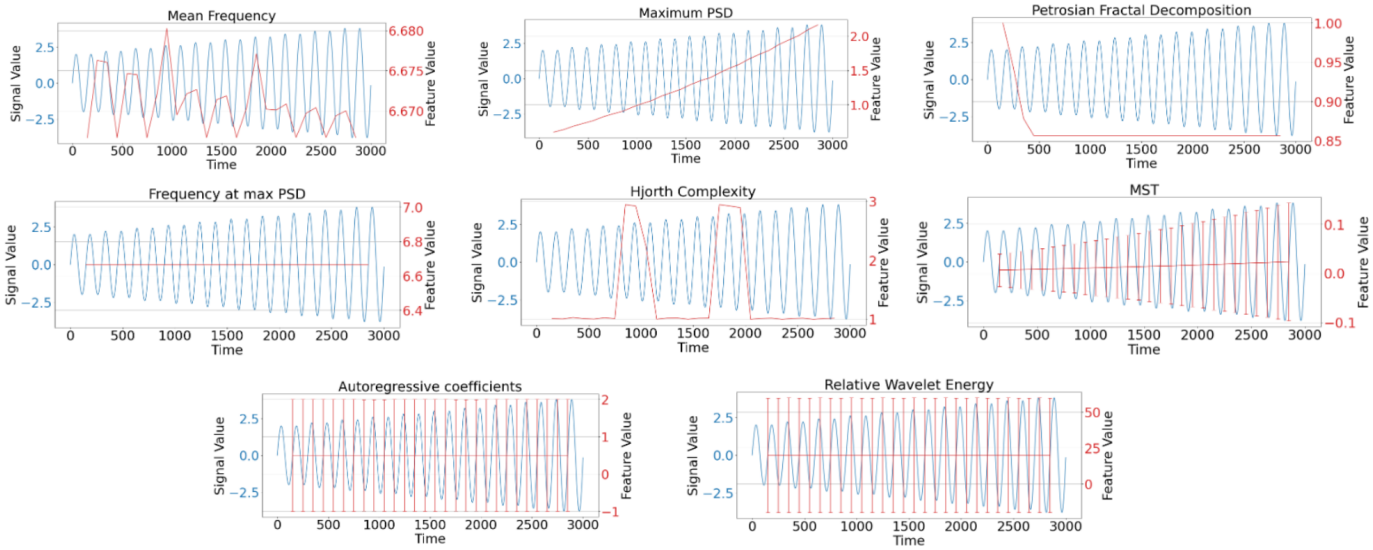

Fig. 4: Some features depend on the amplitude also. These have been represented on a concatenated sine wave of different amplitudes but of the same frequency. Each sine wave was of length 300, and the features were calculated for each sine wave

with changing frequency and amplitude (Fig.3, Fig.4).

8) Mean frequency feature vector - The mean frequency of the Fourier transformed data was used. The feature value increases with frequency(Fig. 3) but shows a complex relationship with Amplitude.

D. Feature Selection

Feeding all the features would lead to higher dimensional feature vectors and hence more trainable parameters of the machine learning classifier. Features capturing redundant and irrelevant features makes training robust classifiers challenging. Hence, a genetic algorithm feature selection mechanism is used to select the optimal combination of features intelligently.

The genetic algorithm randomly initializes multiple sequences of 0s and 1s, equivalent to the length of the feature vector. 1 means the feature is selected, and 0 means it is not. This array of 0s and 1s is called chromosomes. Multiple chromosomes consist of a population. The training data's five-fold cross-validation accuracy evaluates each chromosome's fitness. Mutations and crossovers happen between a population

Feature Set	Dataset	Maximum accuracy before signal Split (%)	Maximum accuracy after signal split (%)	Optimum Signal split
MST	BCI3	86	86	1
	GC_MOT	65.27	76.03	3
	JC_MOT	70.2	89.3	2
	JT_MOT	43	57.8	3
	WC_MOT	67.7	85.07	2
	RR_MOT	83.47	98.33	10
RWE	BCI3	83	86	2
	GC_MOT	67.73	71.03	6
	JC_MOT	78.4	84.23	3
	JT_MOT	48.03	50.47	10
	WC_MOT	74.33	85.07	3
	RR_MOT	86.83	93.43	3
AR	BCI3	83	86	2
	GC_MOT	67.73	71.03	6
	JC_MOT	78.4	84.23	3
	JT_MOT	48.03	50.47	10
	WC_MOT	74.33	85.07	3
	RR_MOT	86.83	93.43	3
PFD	BCI3	84	89	2
	GC_MOT	56.83	66	2
	JC_MOT	52.07	71.87	6
	JT_MOT	57	58.63	2
	WC_MOT	68.67	84.27	10
	RR_MOT	67.77	79.33	6
MFR	BCI3	82	82	1
	GC_MOT	67.77	68.6	2
	JC_MOT	79.3	84.3	2
	JT_MOT	47.2	54.5	3
	WC_MOT	74.3	82.63	2
	RR_MOT	88.47	95.03	3
FMP	BCI3	74	83	3
	GC_MOT	50.47	59.4	2
	JC_MOT	56.07	66.87	3
	JT_MOT	51.23	51.23	1
	WC_MOT	66.9	66.9	1
	RR_MOT	74.37	92.63	10
PSD	BCI3	69	82	30
	GC_MOT	60.23	72.7	2
	JC_MOT	72.63	80.27	2
	JT_MOT	48	51.23	10
	WC_MOT	74.3	74.33	10
	RR_MOT	82.73	90.17	10
HC	BCI3	54	61	2
	GC_MOT	50.4	51.23	30
	JC_MOT	50.4	52.9	15
	JT_MOT	51.23	51.23	1
	WC_MOT	50.4	52.9	2
	RR_MOT	49.57	57.1	30

TABLE I: Individual Feature analysis: Accuracy for each feature set in the BCI3 and motor basic datasets, before and after the signal split analysis

to generate new and fitter chromosomes. After a specified number of generations, the fittest chromosome is chosen, and hence a subset of features based on that chromosome is selected.

E. Classification

Machine learning classifiers directly enable learning patterns from the example data without relying on predefined equations as a model [36]. The Machine learning classification algorithm used in this study is the Support vector machine (SVM). SVM creates a hyperplane (or a line for only two features) or a decision boundary around the data points to classify a new data point [37].

F. Information Transfer Rate Calculation

The ITR is calculated according to the following equations:

$$B = \log_2 N + P \log_2 P + (1-P)\log_2(\frac{1-P}{N-1}) \quad (1)$$

$$ITR(\frac{Bits}{min}) = B \times 60/T \quad (2)$$

Where N is the number of classes and P is the classification accuracy and T is the trial length in seconds.

III. Results and Discussion

A. Individual Feature Analysis

Multiple features were extracted from one electrode's signal in the feature sets of MST, RWE, and AR. For example, there were 5 features per channel (electrode) of the Relative Wavelet Energy feature set, and as there are 64 electrodes, there were 64*5 = 320 features in each trial for the Relative Wavelet Energy feature set.

Xu et al. [6], Zhao et al. [7], and Hammon et al. [8] respectively managed to achieve 98%, 91.8%, and 87% using MST, Relative Wavelet Energy, and Auto-regressive coefficients (along with spectral power estimate features). The difference in individual features for BCI3 in our case (Table I) and the above-mentioned studies is due to various factors, including differences in preprocessing schemes, Filtering, Dimensionality reduction, Feature parameters, Channel Selection, and Classifiers and Classifier Hyperparameters.

The individual feature accuracies set a baseline to compare with signal split feature analysis. In signal split, the signal is split into several sub-segments; for example, a signal split of 2 means the signal of length 3000 is split into two sub-segments, each of length 1500. The features are extracted on each of the sub-segments and then concatenated. As demonstrated in Fig. 3, there is significant variation in the feature values as the frequency and amplitude of the sine wave change, and the changing amplitude and frequency of the sine wave replicate a real-world signal where amplitude and frequency are continuously changing. So if we take a single value for the whole signal, it would not represent it as accurately. Thus, we use the signal split parameter, which divides the signal into n sub-signals on which the value of the feature is calculated, and thus the non-stationary nature of the signal is accounted for.

In Table I, we see that the accuracy increases in 43 out of 48 cases across the eight feature sets and six subjects. It can also be observed that different feature sets have different optimal splits. For feature sets MST, AR, MFR, and FMP, the average optimal split is between 2 and 3.5, which means these feature sets give a better result when extracted on a long signal length, so these can be called long-range features. The feature sets RWE and PFD had an average optimal split of 4.5 and 4.67.

Subject	Total Features Per Trial	Maximum accuracy before Feature Selection (%)	Maximum accuracy after Feature Selection (%)
BCI3	3136	90	92
GC_MOT	3136	70.2	91.76
JC_MOT	2352	83.36	90.17
WC_MOT	3136	77.6	92.6
JT_MOT	3038	47.87	90.93
RR_MOT	2401	90.1	89.3

TABLE II: Feature Selection analysis: The accuracy for the subject before and after the feature selection. The total number of features per trial are the feature values that were extracted for all eight feature sets and all electrodes and then combined.

These feature sets give better results when used for a moderate signal length. Thus, these are called mid-range features. The feature sets of PSD and HC had an average optimal split of 10.67 and 13.33, these feature sets would be most useful when used for short lengths, so these are short-range features. Thus, a signal split parameter should also be optimized whenever individual features are used to achieve the best results.

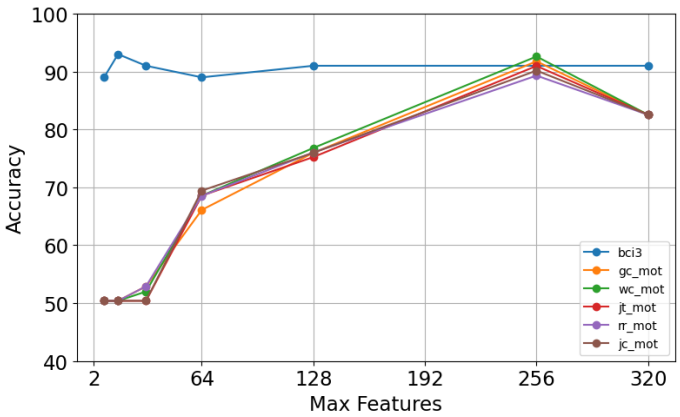

Fig. 5: Accuracy vs Max Features for BCI3 and all motor basic dataset subjects. The maximum number of features the genetic algorithm could select is the maximum.

B. Feature selection

Although in the case of MST feature for BCI3, the accuracy is 86% it consists of 2240 (MST extracts 35 features from each electrode, and there are 64 electrodes hence 35*64=2240 features) features, and according to Table II, the accuracy when all the features are combined is 90% whereas genetic selection achieves 93% using 16 max features, max features is an arbitrary parameter which restricts the number of features the genetic algorithm can select, the reduced number of features results in faster classification [6]. As compared to state of the art for BCI3 dataset given in Table III this approach achieves high accuracy with a low dimensional feature vector.

Similarly, as in the BCI3 dataset, the accuracy increases significantly in 5 out of the 6 subjects, and in one subject, the accuracy only slightly decreases by 0.8%. As shown in Fig. 5, the maximum accuracy is obtained by the genetic algorithm at 16 max features for the BCI 3 dataset and 256 max features for motor basic subjects, which is significantly less than the total number of features. This means that the genetic selection method can obtain an optimal combination of features that complement each other. A similar approach was used by Wei et al. (2010) [38] and Wei et al. (2008) [39]. They used the genetic algorithm to select the best combination of channels and then extracted features from the selected channels. Although, the difference from our approach is that we selected the features after extraction of these features from all of the channels, as our objective was to create a pool of selected features that best complement each other.

Reference	Features Used	Accuracy	Length of Final Feature Vector	Channel Selection Method
Xu et al., 2020 [6]	MST	98	455	Wrapper Optimization
Xu et al., 2016 [9]	Fractal, Local Binary Pattern	95	3321	Cross Validation
Zheng et al., 2019 [21]	MST	95	1260	Cross Validation
Yan and Guan, 2013 [40]	Wavelet Variance	88	12	Wavelet Variance
Zhao et al., 2010 [7]	RWE	84	40	Visual Inspection
Present Work	**MST, RWE, AR, PFD, MFR, FMP, PSD, HC**	**93**	**16**	**Genetic Algorithm**

TABLE III: State of the art for the BCI3 dataset

Fig.6 shows the number of selected features in each feature set after genetic selection with maximum features as 16, 32, 64, 128, and 256. It shows how the genetic algorithm selects the features in different ratios for the perfect combination. It shows how the genetic algorithm selects the features in different ratios for the perfect combination. Any feature from the feature set HC is not selected except when 256 max features are used. This means that the genetic algorithm filters out any feature from the Hjorth Complexity feature set as it determines that it is not in the best 128 features.

C. Information Transfer Rate Analysis

The subjects in both datasets were told to imagine the movement when the data recording started. Hence, it is unclear if the whole 3 seconds of data are required for classification, so we wanted to see if we could achieve comparable accuracy with the reduced signal length. The reduced signal length would result in a higher ITR. The signal from the first 500, 1000, 1500, 2000, 2500, or 3000ms was used to compute the features, and the ITR was calculated.

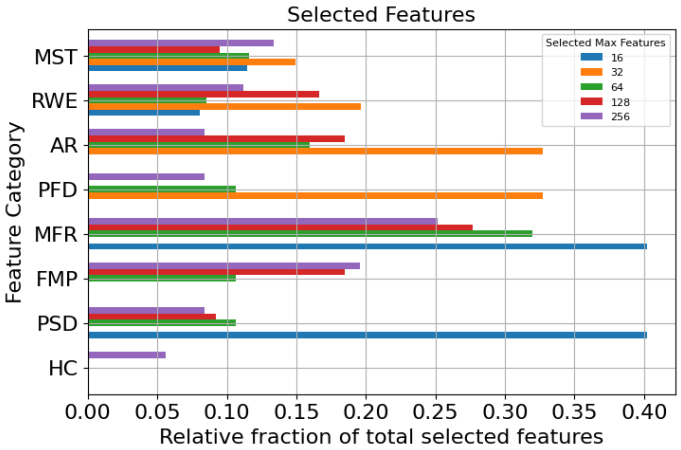

Fig. 6: The number of features selected in the BCI3 dataset from each of the eight feature sets. The number of features from each feature set is different, so to give the relative fraction of a set, the selected features from a feature set were normalized by dividing by the number of features in that particular set.

Subject	ITR increase (Bits/min)	Length of Signal at maximum ITR (%)	Classification Accuracy Reduction at maximum ITR (%)
BCI3	8.58	1000	9.0
GC_MOT	14.72	500	26.56
JC_MOT	29.64	500	20.6
WC_MOT	6.59	1000	19.07
JT_MOT	0	3000	0
RR_MOT	63.72	500	8.23

TABLE IV: ITR Analysis: The increase in ITR after reduced signal length for all of the subjects

We observed that in the case of the motor execution datasets, the ITR has increased for four subjects out of five although there is a decrease in accuracy (Table IV).

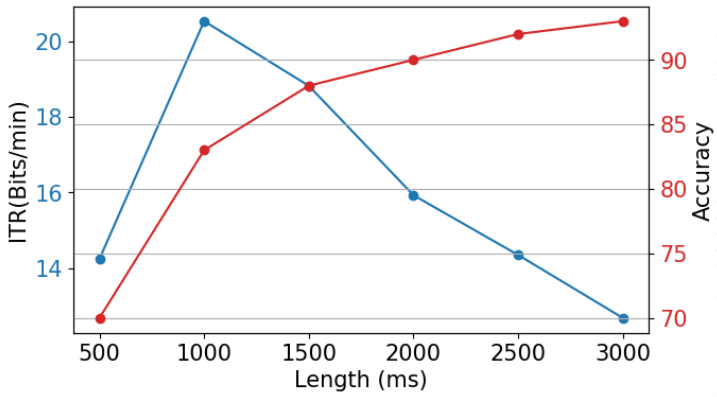

Fig. 7: Accuracy and ITR vs. Length for the BCI3 dataset

In the case of the motor imagery (BCI3) dataset, there is an increase in ITR with a decrease in accuracy of 9% (Table IV). The trend of accuracy and ITR with respect to trial length is shown in Fig. 7. We obtained an ITR of 20.5 for the BCI3 dataset, which is higher than state-of-the-art (Table V). Although there is a decrease of 9% in accuracy for obtaining maximum ITR, we observed that the accuracy decreases only by 1% when the trial length is reduced from 3000 to 2500 and the ITR increases by 2.40 Bits/min (Fig. 7). This suggests that the optimal accuracy-ITR trade-off in the EcoG BCI can be optimized by reducing the signal length.

Reference	ITR (Bits/min)
Xu et al., 2020 [6]	17.11
Xu et al., 2016 [9]	14.27
Zheng et al., 2019 [21]	14.27
Zhao et al., 2010 [7]	11.81
Yan and Guan, 2013 [40]	9.41
Present Work	**20.50**

TABLE V: ITR for state of the art for the BCI3 dataset

IV. Conclusion

Brain-computer interfaces require efficient classification techniques. Various approaches for efficient classification exist, including deep learning and machine learning techniques. Machine learning techniques generally focus on a particular feature extraction technique. Our goal in this study was to enhance the accuracy while using a particular feature by introducing a signal split parameter. As it is difficult to anticipate which feature set would be optimal for a particular task, we encourage using a large feature pool and genetic selection to find an optimal combination of features. The genetic algorithm is time-consuming to train, but once the optimal combination is found, the classification becomes faster and more accurate as the dimensionality of the data is reduced. Due to the reduced dimensionality, the genetic algorithm is also a hardware-friendly technique. Using a reduced signal length along with genetic selection increases the ITR. Hence this approach can be used to design fast and accurate EcoG BCIs.

The current state of the art for BCI3 includes deep learning approaches which use LSTM or CNN for classification [41], [42]. Although more accurate, these methods require high processing power and are generally subject to the BlackBox problem which makes it harder for programmers to understand the functioning of the algorithms and makes it difficult to tweak them. This paper focuses mainly on Machine Learning and Feature extraction methods.

A major limitation of our work is the absence of deep-learning approaches, which is out of this work. Although, the major problem with using Deep Learning approaches with EcoG is the lack of data and the black box problem. Hence, we suggest that end-to-end deep learning methods may not be the best solution. Rather, some feature extraction might be optimal.

V. Appendix

References

[1] DL Keene, S Whiting, and ECG Ventureyra. Electrocorticography. *Epileptic Disorders*, 2(1):57–64, 2000.

[2] David Roberts, Michael Higgin, Enrique Ventureyra, et al. Alfentanil mediated activation of epileptiform activity in the electrocorticogram during resection of epileptogenic foci. *Canadian journal of neurological sciences*, 24(1):29–36, 1997.

[3] Richard A Andersen, Sam Musallam, and Bijan Pesaran. Selecting the signals for a brain–machine interface. *Current opinion in neurobiology*, 14(6):720–726, 2004.

[4] Aysegul Gunduz and Gerwin Schalk. Ecog-based bcis. *Brain–Computer Interfaces Handbook*, pages 297–322, 2018.

[5] Dennis J McFarland, William A Sarnacki, and Jonathan R Wolpaw. Brain–computer interface (bci) operation: optimizing information transfer rates. *Biological psychology*, 63(3):237–251, 2003.

[6] Fangzhou Xu, Wenfeng Zheng, Dongri Shan, Qi Yuan, and Weidong Zhou. Decoding spectro-temporal representation for motor imagery recognition using ecog-based brain-computer interfaces. *JIN*, 19(2):259–272, 2020.

[7] Hai-bin Zhao, Chun-yang Yu, Chong Liu, and Hong Wang. Ecog-based brain-computer interface using relative wavelet energy and probabilistic neural network. In *2010 3rd International Conference on Biomedical Engineering and Informatics*, volume 2, pages 873–877. IEEE, 2010.

[8] Paul S Hammon and Virginia R de Sa. Preprocessing and meta-classification for brain-computer interfaces. *IEEE Transactions on Biomedical Engineering*, 54(3):518–525, 2007.

[9] Fangzhou Xu, Weidong Zhou, Yilin Zhen, Qi Yuan, and Qi Wu. Using fractal and local binary pattern features for classification of ecog motor imagery tasks obtained from the right brain hemisphere. *International Journal of Neural Systems*, 26(06):1650022, 2016.

[10] Lin Yao and Mahsa Shoaran. Enhanced classification of individual finger movements with ecog. In *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, pages 2063–2066. IEEE, 2019.

[11] Qingguo Wei, Xiaorong Gao, and Shangkai Gao. Feature extraction and subset selection for classifying single-trial ecog during motor imagery. In *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1589–1592. IEEE, 2006.

[12] K Aswinseshadri and V Thulasi Bai. Feature selection in brain computer interface using genetics method. In *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*, pages 270–275. IEEE, 2015.

[13] Thomas Lal, Thilo Hinterberger, Guido Widman, Michael Schröder, N Hill, Wolfgang Rosenstiel, Christian Elger, Niels Birbaumer, and Bernhard Schölkopf. Methods towards invasive human brain computer interfaces. *Advances in neural information processing systems*, 17, 2004.

[14] Kai J Miller. A library of human electrocorticographic data and analyses. *Nature human behaviour*, 3(11):1225–1235, 2019.

[15] Steven D Colan. The why and how of z scores. *Journal of the American Society of Echocardiography*, 26(1):38–40, 2013.

[16] Arthur Petrosian. Kolmogorov complexity of finite sequences and recognition of different preictal eeg patterns. In *Proceedings eighth IEEE symposium on computer-based medical systems*, pages 212–217. IEEE, 1995.

[17] Silvia Maria Alessio. Digital signal processing and spectral analysis for scientists: concepts and applications. 2015.

[18] Bo Hjorth. Eeg analysis based on time domain properties. *Electroencephalography and clinical neurophysiology*, 29(3):306–310, 1970.

[19] Sylvain Baillet, Karl Friston, and Robert Oostenveld. Academic software applications for electromagnetic brain mapping using meg and eeg, 2011.

[20] Sari Saba-Sadiya, Eric Chantland, Tuka Alhanai, Taosheng Liu, and Mohammad Mahdi Ghassemi. Unsupervised eeg artifact detection and correction. *Frontiers in Digital Health*, 2:57, 2020.

[21] Wenfeng Zheng, Fangzhou Xu, Minglei Shu, Yingchun Zhang, Qi Yuan, Jian Lian, and Yuanjie Zheng. Classification of motor imagery electrocorticogram signals for brain-computer interface. In *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 530–533. IEEE, 2019.

[22] F Xu, W Zhou, Y Zhen, and Q Yuan. Classification of ecog with modified s-transform for brain-computer interface. *Journal of Computational Information Systems*, 10(18):8029–8041, 2014.

[23] Hongli Chang and Jimin Yang. Genetic-based feature selection for efficient motion imaging of a brain–computer interface framework. *Journal of neural engineering*, 15(5):056020, 2018.

[24] Said Assous and Boualem Boashash. Evaluation of the modified s-transform for time-frequency synchrony analysis and source localisation. *EURASIP Journal on Advances in Signal Processing*, 2012(1):1–18, 2012.

[25] Karl J Friston. Statistical parametric mapping. In *Neuroscience databases*, pages 237–250. Springer, 2003.

[26] Rajkumar Palaniappan, Kenneth Sundaraj, Sebastian Sundaraj, N Huliraj, SS Revadi, and B Archana. Pulmonary acoustic signal classification using autoregressive coefficients and k-nearest neighbor. In *Applied Mechanics and Materials*, volume 591, pages 211–214. Trans Tech Publ, 2014.

[27] BS Raghavendra and D Narayana Dutt. Signal characterization using fractal dimension. *Fractals*, 18(03):287–292, 2010.

[28] Luis Alfredo Moctezuma and Marta Molinas. Classification of low-density eeg for epileptic seizures by energy and fractal features based on emd. *Journal of biomedical research*, 34(3):180, 2020.

[29] Wei Bin Ng, A Saidatul, YF Chong, and Z Ibrahim. Psd-based features extraction for eeg signal during typing task. In *IOP Conference Series: Materials Science and Engineering*, volume 557, page 012032. IOP Publishing, 2019.

[30] Zhi Ying Ong, A Saidatul, and Z Ibrahim. Power spectral density analysis for human eeg-based biometric identification. In *2018 International Conference on Computational Approach in Smart Systems Design and Applications (ICASSDA)*, pages 1–6. IEEE, 2018.

[31] Lin Yao, Jonathan L Baker, Nicholas D Schiff, Keith P Purpura, and Mahsa Shoaran. Predicting task performance from biomarkers of mental fatigue in global brain activity. *Journal of neural engineering*, 18(3):036001, 2021.

[32] Ömer Türk, Mesut Şeker, Veysi Akpolat, and Mehmet Siraç Özerdem. Classification of mental task eeg records using hjorth parameters. In *2017 25th Signal Processing and Communications Applications Conference (SIU)*, pages 1–4. IEEE, 2017.

[33] Seung-Hyeon Oh, Yu-Ri Lee, and Hyoung-Nam Kim. A novel eeg feature extraction method using hjorth parameter. *International Journal of Electronics and Electrical Engineering*, 2(2):106–110, 2014.

[34] Rami J Oweis and Enas W Abdulhay. Seizure classification in eeg signals utilizing hilbert-huang transform. *Biomedical engineering online*, 10(1):1–15, 2011.

[35] Mesut Şeker and Mehmet Siraç Özerdem. Classification of eeg data sets with hilbert transform. In *2016 24th Signal Processing and Communication Application Conference (SIU)*, pages 1949–1952. IEEE, 2016.

[36] Jaime G Carbonell, Ryszard S Michalski, and Tom M Mitchell. An overview of machine learning. *Machine learning*, pages 3–23, 1983.

[37] William S Noble. What is a support vector machine? *Nature biotechnology*, 24(12):1565–1567, 2006.

[38] Qingguo Wei, Zongwu Lu, Kui Chen, and Yuhui Ma. Channel selection for optimizing feature extraction in an electrocorticogram-based brain-computer interface. Technical Report 5, 2010.

[39] Qingguo Wei and Wei Tu. Channel selection by genetic algorithms for classifying single-trial ecog during motor imagery. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 624–627. IEEE, 2008.

[40] Shi Yu Yan and De Jun Guan. Ecog classification research based on wavelet variance and probabilistic neural network. In *Applied Mechanics and Materials*, volume 380, pages 2280–2285. Trans Tech Publ, 2013.

[41] Mamunur Rashid, Minarul Islam, Norizam Sulaiman, Bifta Sama Bari, Ripon Kumar Saha, and Md Jahid Hasan. Electrocorticography based motor imagery movements classification using long short-term memory (lstm) based on deep learning approach. *SN Applied Sciences*, 2(2):1–7, 2020.

[42] Fangzhou Xu, Fenqi Rong, Yunjing Miao, Yanan Sun, Gege Dong, Han Li, Jincheng Li, Yuandong Wang, and Jiancai Leng. Representation learning for motor imagery recognition with deep neural network. *Electronics (Switzerland)*, 10(2):1–13, jan 2021.