EEG Classification based on Machine Learning Techniques

Farid Ali Mousa
Assistant Lecturer
Faculty of Computers and Information,
Beni Suef University

Reda A. El-Khoribi
Professor
Faculty of Computer and Information,
Cairo University

Mahmoud E. Shoman
Professor
Faculty of Computer and Information,
Cairo University

ABSTRACT
The main issue to build applicable Brain-Computer Interfaces is the capability to classify the electroencephalograms (EEG). During the last decade, researchers developed lots of interests in this field. The purpose behind this research is to improve a model for EEG signals analysis. Filtration of EEG Signals is essential to remove artifacts. Otherwise, wavelet transform was used to extract features. Mean, Maximum, Minimum and Standard Deviations values of wavelet coefficients for the EEG signals were chosen as a feature vector. This paper compares the classification results by the use of Neural Network, K-Nearest Neighbor and Support Vector Machine classifiers. It has been illustrated from results that the K-Nearest Neighbor classifier outperforms a better performance than Neural Network and Support Vector Machine.

Keywords
Brain Computer Interface; Support Vector Machine; Neural Network; K-Nearest Neighbor; Wavelet Transform; EEG.

1. INTRODUCTION
The system that allows the brain signals for the environmental interaction is called The Brain Computer Interface (BCI). This system has been divided into two groups; the first one is called invasive while the other one is non-invasive. [1] The differences between the two groups are the invasive devices are attached directly to the brain and their signals are with high quality. On the other hand, the Non-invasive clarity is very low when it communicates with the brain. Moreover, it considered to be very safest in comparison to other types.BCI is a device that receives neural signals and transforms them to digital signals that a computer can utilize them for a lot of purposes. This device can be extremely essential and powerful as it can act like a bridge to connect the gap between the human body and the tools used for the environmental interaction. Instead of being forced to manipulate a real or symbolic environment with body parts like hands and legs, users should be able to interact with their environment using the brain's neural impulses. [2]

IN 1875 EEG signals were recorded as a brain activity by Richard Caton (1842-1926), who placed a pair of galvanometer electrodes on the surface of the scalp. His work thus followed Matteucci Carl (1811-1868) and Emil Du Bois-Reymond (1818-1896) that is managed to use a galvanometer to measure neural activity of the muscle. This field of medicine known as neurophysiology [3].One of the most significant factors to measure the abnormalities in clinical EEGs and to understand the useful behaviors in cognitive research is the frequency. The human EEG are expressed as temporary unexpected alterations with irregular bursts of oscillations by the use of too many (billions) pulses from neurons. These signals are classified into five bands namely, delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz) and gamma (> 30 Hz).This classification is vital for the brain diseases diagnosis also for better understanding the cognitive processes. While to differentiate EEG segments and for taking a decision regarding the human health, effective classification techniques were used. Although EEG recordings characterized by having many data, but still a main problem to use the documented EEG signals for additional analytical process (e.g. classification). The advantages behind this is to separate the useful features alone from the raw EEG signals after this separation, these extracted features can be used for the classification. Thus a model is developed to make this extraction and for comparing the results by the use of KNN, SVM and Neural Network.Several methods have been discussed in the literature to classify these signals such as neural networks [4], statistical methods [5], Adaptive neuro-fuzzy inference system [6], many other studies [7], [8], [9] that covered the brain signals analysis.For EEG classification, many methods have been proposed in the literature, namely, neural networks [4], statistical methods [5], Adaptive neuro-fuzzy inference system [6], also a number of other studies [7], [8], and [9] that covered analyzing brain signals.The remaining part of the paper is organized as follows. Preliminaries are discussed in sector (2). Sector (3) presents the proposed model. Experimental results are discussed in sector (4). Finally, the paper conclusion is in sector (5).

2. PRELIMINARIES

2.1 Wavelet transform
The wavelet transform is a method that used to decompose an input signal of interest into a set of coefficients and provides a way to analyze the signal by examining these coefficients, many methods have been proposed in the literature that used wavelet transform for feature extraction [10], [11], [12], [13], and [14].The extracted wavelet coefficients give compacted representation that shows the distribution of the EEG energy in time and frequency domain.To obtain the decomposition of the signal into different frequency bands, two main filtering process are done namely high-pass and low-pass. The decomposed bands are called sub bands. Both filters provide different types of the coefficients as an output. The low-pass filter provides the approximation coefficients. On the other hand, the high pass filter provides the detailed one. The analysis of the signals using DWT depends on selecting an appropriate wavelet and the number of levels of decomposition is the main coefficients. The number of levels of decomposition is chosen based on the dominant frequency components of the signal. The levels are chosen such that those parts of the

signal that correlate well with the frequencies required for classification of the signal are taken in the wavelet coefficients.

Algorithm 1 " Wavelet Transform

S is a symbol given for the signal of length N
Level from j=1 to maximum level

1: two sets of coefficients were produced: Approximation coefficients cA1, and detail coefficients cD1.

2: These vectors are obtained by convolving s with the low-pass filter Lo_D for approximation, and with the high-pass filter Hi_D for detail

2.1: A coefficients

$$A_j = \sum_k A_k^{(j)} \emptyset_{j,k} \quad and\ A_{j+1} = \sum_k A_k^{(j+1)} \emptyset_{j+1,k}$$

$$A_k^{(j+1)} = \sum_n h_{n-2k}\, A_n^j$$

$$\tilde{h}(k) = h(-k), and\ F_k^{(j+1)} = \sum_n \tilde{h}_{k-n}\, An^j$$

2.2: D coefficients

$$D_1 = \sum_n \delta_n \varphi_{1,n}$$

Such that $\varphi_{j+1,0} = \sum g_k\, \emptyset_{j,k}$

2.2 Artificial Neural Network (ANN)

We use Artificial Neural Network (ANN) technique in classification step because it uses to solve complexity problems. Artificial network adapts itself by sequential training algorithm and its architecture and connected weights. This paper uses feed forward neural network with multi-layers with back propagation learning algorithm.

Algorithm 1 Artificial Neural Network (ANN) algorithm

1: initialize weights (set to small random value).

while stopping condition is false do steps 2-9

2: for each sample in the training set, do steps 3-8

Feed forward :-

3: Each input unit (Xi) receives signal Xi & broad casts this signal to all units in the layer above (the hidden layer)

4: Each hidden unit (Zj) sums its weighted i/p signals,

$$Z - i_{nj} = V_{aj} + \sum_{i=1}^{n} x_i v_{ij} \quad , Where\ V_{aj}\ is\ a\ bias$$

To compute its output signal, apply its activation function

$$Z_j = 1/(1 + e^{-(Z-inj)})$$

send this signal to all units that present in the layer above

5: Calculate the output:

$$Y - i_{nk} = W_{ok} + \sum_{i=1}^{n} Z_j w_{jk} \quad , Where\ W_{ok}\ is\ a\ bias$$

$$Y_k = 1/(1 + e^{-(Y-i_{nk})})$$

Back propagation of error:-

6: Computes the error in the output layer

$$\delta_{2k} = Y_k(1 - Y_k) * (T_k - Y_k), T_k\ is\ the\ target$$

7: computes its error information in hidden layers

$$\delta_{1j} = Z_j\big(1 - Z_j\big) * \sum_{k=1}^{m} \delta_{2k}\, w_{lk},$$

8: Update weights and bias :-

$$W_{jk}(new) = \eta * \delta_{2k} * Z_j + \alpha * W_{jk}(old)$$

$$V_{ij}(new) = \eta * \delta_{1j} * x_i + \alpha * V_{ij}\, old$$

9: Test stopping condition.

2.3 K-Nearest Neighbor Classifier (KNN)

In pattern recognition, one of the methods used to classify the objects is the k-nearest neighbor algorithm (KNN). This method is used depending on training examples close to each other in the feature space and it is used in many applications (e.g. medical field, data mining, recognition of handwriting, statistical pattern recognition, and satellite image).

Algorithm2" K-Nearest Neighbor Classifier (KNN) algorithm"

1: Repeat the following steps to all samples

2. Calculate the distance between sample x and all samples in the training data

$$D = \sqrt[2]{\sum (X_i - S_i)^2}$$

3: Sort the distances ascending

4: Pick first K elements

5: $Output_X = majority_{Class}\,(elements)$

2.4 Support Vector Machine (SVM)

For classification, SVM is a popular machine learning algorithm that is commonly used and it is used for optimizing the margin between two classes. The classification is achieved by understanding a linear or non-linear separation surface in the input space. An important property of SVMs is their ability to learn can be independent of the dimensionality of the feature space.

In the support vector machine algorithm the complexity of the optimization problem is mainly based on the margin with which separate the data, not with the number of features. SVMs use over fitting protection to handle the large feature spaces. SVMs have a lot of advantages and one of them is its ability to learn different kernel function for classification. IN their basic form, SVMs learn the linear threshold function. However, by changing to an appropriate kernel function, they can be used to learn polynomial classifiers, radial basic function (RBF) networks and three-layer sigmoid nets. [15]

Algorithm 3 " Support Vector Machine (SVM) algorithm"

1: Finding Pair of Points that are closed

candidateSV ={closest pair from classes that are opposite}

do

Find a violator

2: Adding a sample to the Support Vector data set

candidateSV = candidateSV ∪ violator

3: Pruning

if any $\alpha_p < 0$ *as a result of adding c to S then*

candidateSV = candidateSV \ p

repeat till all such points are pruned

end if

while there are violating points do

3. PROPOSED ALGORITHM

3.1 Dataset

The data set used in this experiment has been obtained from the BCI Competition IV (2008). Provided by Department of Medical Informatics, Institute for Biomedical Engineering, University of Technology Graz [16], the data set consists of EEG data recorded from 9 subjects. For a given subject two sessions were recorded on two different days. Each session consisted of 6 runs and each run consisted of 48 trials (12 trials for each motor imagery class). There are four different motor imagery tasks, namely the imagination of movement, Left hand (class 1), Right hand (class 2), Both feet (class 3), Tongue (class 4).

All data sets are stored in the General Data Format for biomedical signals (GDF), One file per subject and session, only one session contains the class labels for all trials, whereas the other session will be used to test the classifier and hence to evaluate the performance. For each subject we have 25 channels (22 EEG and 3 EOG).

At the beginning of each session, a recording of approximately 5 minutes was performed to estimate the EOG influence. The recording was divided into 3 blocks:

(1) two minutes with eyes open (looking at a fixation cross on the screen),

(2) one minute with eyes closed, and

(3) one minute with eye movements.

3.2 Methods

In this section we show the details for a proposed method to classify EEG signals using machine learning techniques.

A) Preprocessing

After reading the GDF files; it's needed to remove unwanted signals present in the EEG. They have various origins, which include the utility frequency, body and eye movements, or blinks. The utility frequency artifacts were already removed from the data using the notch filter; the step of eye artifacts removal is processed here using high pass filter because the eye artifacts occur in the frequency range of 0–4 Hz, so by filtering these components out, we may reduce the EOG artifacts.

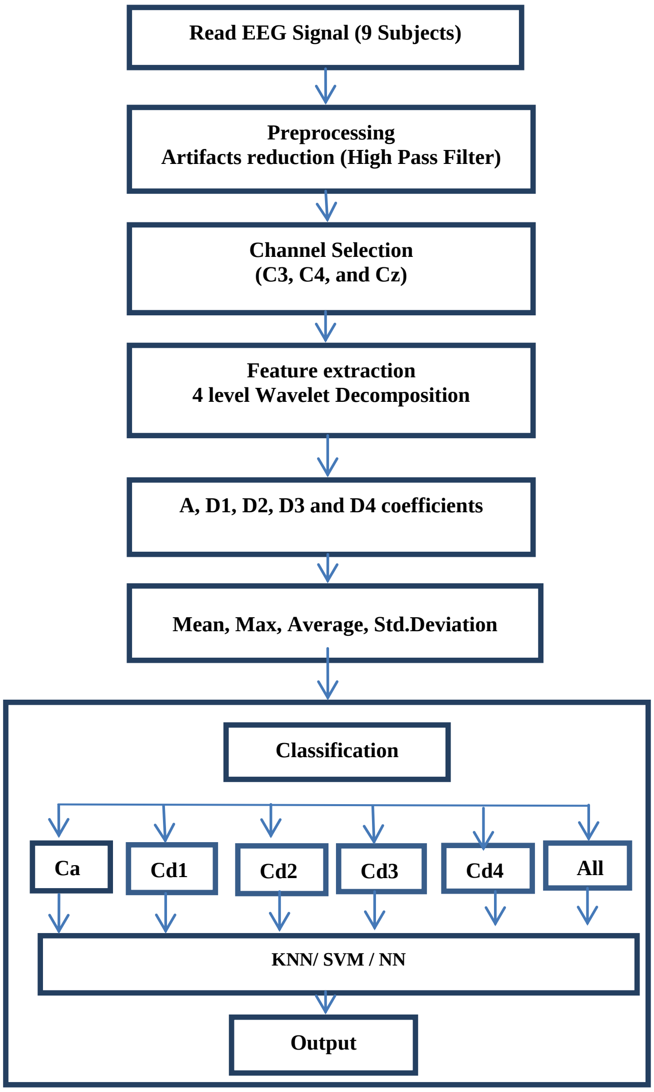

Fig 1: Proposed System

B) Channel selection

Only channels C3, C4, and Cz were selected for analyzing EEG signals because: (1) It is stated in [17] that most EEG channels represent redundant information and (2), it was concluded in[18], [19] that the neural activity that is mostly correlated to the fists movements is commonly contained within these channels.

C) Feature extraction

In this experiment the signal has been decomposed in 4 levels using discrete wavelet transform. The approximate coefficient Ca and each level detail coefficients Cd4, Cd3, Cd2 and Cd1 were used to get the feature vector. Many amplitude estimators for neurological activities were defined mathematically to get the feature vector in this paper we use minimum, maximum, Average and standard deviation. If we assume that the *nth* sample of a wavelet decomposed detail at level i is Di (n), then the following features can be defined:

1. Minimum

$$Minimum = Min\{D_i\}$$

2. Maximum

$$Maximum = Max\{D_i\}$$

3. Average

$$\mu = \frac{1}{N}\sum_{j=1}^{N} D_j$$

4. Standard Deviation

$$\sigma = \sqrt{\frac{1}{N}\sum (x_i - \mu)^2}$$

D) Classification

In this paper we classify the data using A coefficients, D_1 coefficients, D_2 coefficients, D_3 coefficients, D_4 coefficients, and finally we classify using a feature vector that composed from A, D_1, D_2, D_3, D_4 coefficients.

4. EXPERIMENTAL RESULTS

In this section we describe the results obtained from classification using Support Vector Machine (SVM), Neural Network (NN) and K-Nearest Neighbor (KNN).

A) Classification using A Coefficients

Table 1. A coefficient Results

Subject	SVM	NN(7,5,1)	KNN, K=7
1	51.4	40.3	50
2	44.4	50	66.7
3	56.9	51.4	69.4
4	52.8	50	69.4
5	48.6	48.6	69.4
6	47.2	41.7	44.4
7	52.8	45.8	70.8
8	48.6	43.1	66.7
9	50	54.1	68.1
Mean Accuracy	**50.3**	**47.2**	**63.8**

As shown in table 1 we can see that KNN classifier take only one parameters and the best results was 63.88% were when k=7.While the SVM classifier with RBF Kernel function the accuracy is 50.3%. In NN we pick number of layers and nodes in each layer to train the samples and get the weights of the network, we test the data in different cases in case of using three layers, in the first layer we pick 7 nodes, 5 nodes in the second layer and 1 nodes in the output layer we got 47.2% accuracy. The accuracy of KNN better than in the SVM and NN classifiers.

B) Classification using D1 Coefficients

Table 2. D1 coefficient Results

Subject	SVM	NN(7,7,1)	KNN, K=35
1	54.2	58.9	72.2
2	43.1	62.5	70.8
3	41.7	59.7	61.1
4	40.3	59.7	62.5
5	38.9	54.2	62.5
6	33.3	43.1	58.3
7	50	58.3	73.6
8	50	58.3	72.2
9	52.8	61.1	59.7
Mean Accuracy	**44.9**	**57.3**	**66.2**

As shown in table 2 we can see that KNN classifier take only one parameters and the best results was 66.2% were when k=35.While the SVM classifier with RBF Kernel function the accuracy is 44.9%. In NN we pick number of layers and nodes in each layer to train the samples and get the weights of the network, we test the data in different cases in case of using three layers, in the first layer we pick 7 nodes, 7 nodes in the second layer and 1 nodes in the output layer we got 57.3% accuracy. The accuracy of KNN better than in the SVM and NN classifiers.

C) Classification using D2 Coefficients

Table 3. D2 coefficient Results

Subject	SVM	NN(7,7,1)	KNN, K=35
1	41.7	50	69.4
2	43.1	51.4	70.8
3	40.3	55.6	69.4
4	38.3	54.2	70.8
5	43.1	59.7	68.1
6	38.3	50	63.9
7	40.3	52.9	70.8
8	43.1	58.3	66.7
9	41.7	50	70.8
Mean Accuracy	**41.1**	**53.6**	**69**

As shown in table 3 we can see that KNN classifier take only one parameters and the best results was 69% were when k=35.While the SVM classifier with RBF Kernel function the accuracy is 41.1%. In NN we pick number of layers and nodes in each layer to train the samples and get the weights of the network, we test the data in different cases in case of using three layers, in the first layer we pick 7 nodes, 2 nodes in the second layer and 1 nodes in the output layer we got 53.56% accuracy. The accuracy of KNN better than in the SVM and NN classifiers.

D) Classification using D3 Coefficients

Table 4. D3 coefficient Results

Subject	SVM	NN(7,7,1)	KNN, K=35
1	34.7	59.7	70.8
2	36.1	58.3	72.2
3	34.7	56.9	70.8
4	37.5	56.9	68.1
5	36.1	58.3	70.8
6	40.2	48.6	66.7
7	36.1	56.9	73.6
8	37.5	58.3	72.2
9	36.1	48.6	72.2
Mean Accuracy	**36.6**	**55.9**	**70.8**

As shown in table 4 we can see that KNN classifier take only one parameters and the best results was 70.8% were when k=35.While the SVM classifier with RBF Kernel function the accuracy is 36.6%. In NN we pick number of layers and nodes in each layer to train the samples and get the weights of the network, we test the data in different cases in case of using three layers, in the first layer we pick 7 nodes, 7 nodes in the second layer and 1 nodes in the output layer we got

55.9% accuracy. The accuracy of KNN better than in the SVM and NN classifiers.

E) Classification using D4 Coefficients

Table 5. D4 coefficient Results

Subject	SVM	NN(7,5,1)	KNN, K=35
1	61.1	62.5	62.5
2	61.1	62.5	58.3
3	65.3	61.1	72.2
4	62.5	62.5	70.8
5	62.5	63.8	61.1
6	54.2	47.2	58.3
7	65.3	56.9	75
8	63.8	56.9	69.4
9	62.5	61.1	72.2
Mean Accuracy	**62**	**59.4**	**66.7**

As shown in table 5 we can see that KNN classifier take only one parameters and the best results was 66.7% were when k=35.While the SVM classifier with RBF Kernel function the accuracy is 62%. In NN we pick number of layers and nodes in each layer to train the samples and get the weights of the network, we test the data in different cases in case of using three layers, in the first layer we pick 7 nodes, 5 nodes in the second layer and 1 nodes in the output layer we got 59.4 % accuracy. The accuracy of KNN better than in the SVM and NN classifiers.

F) Classification using A,D1, D2, D3 and D4 Coefficients

Table 6. All coefficient Results

Subject	SVM	NN(7,5,1)	KNN, K=32
1	43.1	45.8	58.3
2	44.4	51.4	62.5
3	55.6	50	59.7
4	52.8	51.3	63.9
5	50	48.6	61.5
6	51.4	36.1	51.4
7	52.8	52.8	73.6
8	55.6	48.6	72.2
9	56.9	45.9	70.8
Mean Accuracy	**51.4**	**47.8**	**63.8**

As shown in table 6 we can see that KNN classifier take only one parameters and the best results was 63.8% were when k=32.While the SVM classifier with RBF Kernel function the accuracy is 51.4%. In NN we pick number of layers and nodes in each layer to train the samples and get the weights of the network, we test the data in different cases in case of using three layers, in the first layer we pick 7 nodes, 5 nodes in the second layer and 1 nodes in the output layer we got 47.83 % accuracy. The accuracy of KNN better than in the SVM and NN classifiers.

5. CONCLUSIONS

In this study we introduce a model to analyze EEG signals, The proposed technique is based on removing artifacts using High Pass Filter (HPF), and feature extraction using 4 levels discrete wavelet transform; finally we use KNN, SVM and neural network in the classification phase.Good results were obtained using KNN classifier than SVM and NN, for all coefficients, also Maximum classification accuracy was estimated when the statistical parameters of coefficients d3 and d2 were used as the features for classification.

6. REFERENCES

[1] McFarland, Dennis J., and Jonathan R. Wolpaw. "Brain-computer interface operation of robotic and prosthetic devices." Computer 41.10 (2008): 52-56.

[2] BRAIN- COMPUTER INTERFACING WITH EEG: A LOOK AT EYE MOVEMENTS, Jordan E. DeLong, May, 2007

[3] Saeid Sanei, J. A. Chambers. EEG Signal Processing. : John Wiley & Sons, 2008.312p. ISBN 0470511931, 9780470511930.

[4] Qiu W., Fung K. S. M., Chan F. H. Y., Lam F. K., Poon P. W. F., and Hamernik R. P. (2002), ―Adaptive filtering of evoked potentials with Radial-Basis-Functions neural network prefilter,‖ IEEE Trans. Biomed. Eng., vol. 49, 225-212.

[5] Borisoff J. F., Mason S. J., Bashashati A., and Birch G.E.(2002),―Brain-Computer Interface design for asynchronous control applications: improvements to the LF-ASD asynchronous brain switch,‖ IEEE Trans. Biomed. Eng., vol. 51, June,.985-993.

[6] İnan Güler, Elif Derya Übeyli, (2005) ―Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients‖, Journal of Neuroscience Methods, Vol. 148, Issue 2, 30 October, 113-121.

[7] Kai Keng Ang, Zheng Yang Chin, Chuanchu Wang, Cuntai Guan, and Haihong Zhang. Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b. Frontiers in Neuroscience, 6, 2012.

[8] Nicolas Brodu, Fabien Lotte, and Anatole L´ecuyer. Comparative study of band-power extraction techniques for motor imagery classification. In Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB), 2011 IEEE Symposium on, pages 1–6. IEEE, 2011.

[9] Emily M. Mugler, Carolin A. Ruf, Sebastian Halder, Michael Bensch, and Andrea Kubler. Design and implementation of a P300-based brain-computer interface for controlling an internet browser. Neural Systems and Rehabilitation Engineering, IEEE Transactions on, 18(6):599–609, 2010.

[10] Vapnik. The Nature of Statistical Learning Theory. Springer-Verlag, 2000.

[11] A. Glavinovitch, M. N. S. Swamy, and E. I. Plotkin, "Wavelet-Based Segmentation Techniques in the Detection of Microarousals in the Sleep EEG," in 48th Midwest Symposium on Circuits and Systems. 2005, pp. 1302–1305, IEEE.

[12] P. Johankhani, V. Kodogiannis, and K. Revett, "EEG Signal Classifi- cation Using Wavelet Feature Extraction and Neural Networks ," in IEEE John Vincent Atanasoff 2006 International Symposium on Modern Computing (JVA06). 2006, pp. 120–124, IEEE.

[13] C. Dimoulas, G. Kalliris, G. Papanikolaou, and A. Kalampakas, “Long-Term Signal Detection, Segmentation and Summarization Using Wavelets and Fractal Dimension: A Bioacoustics Application in Gastrointestinal-Motility Monitoring,” Comput. Biol. Med., vol. 37, no. 4, pp. 438–462, 2007.

[14] I. W. Selesnick, R. G. Baraniuk, and N. G. Kingsbury, “The Dual-Tree Complex Wavelet Transform,” IEEE Signal Processing Magazine, vol. 22, no. 6, pp. 123–151, 2005.

[15] Nazareth P. Castellanos and Valeri A. Makarov, “Recovering EEG Brain Signals: Artifact Suppression with Wavelet Enhanced Independent Component Analysis,” Journal of Neuroscience Methods, vol. 158, no. 2, pp. 300–312, 2006.

[16] Clemens Brunner, Robert Leeb, Gernot R. M¨uller-Putz, Alois Schl¨ogl, and GertPfurtscheller. BCI Competition 2008 — Graz data set A.

[17] J. Sleight, P. Pillai, and S. Mohan, "Classification of Executed and Imagined Motor Movement EEG Signals," Ann Arbor: University of Michigan, 2009, pp. 1-10.

[18] L. Deecke, H. Weinberg, and P. Brickett, "Magnetic fields of the human brain accompanying voluntary movements: Bereitschaftsmagnetfeld," Experimental Brain Research, vol. 48, pp. 144-148, 1982.

[19] C. Neuper and G. Pfurtscheller, "Evidence for distinct beta resonance frequencies in human EEG related to specific sensorimotor cortical areas," Clinical Neurophysiology, vol. 112, pp. 2084-2097, 2001. Bowman, M., Debray, S. K., and Peterson, L. L. 1993. Reasoning about naming systems.