Motor Imagery Based EEG Signal Classification using Deep Learning Algorithms for BCI Development

Muhammad Zeeshan Baig
2011-NUST-MS-RIME-10
Department of
Robotics & Artificial Intelligence,
School of Mechanical & Manufacturing Engineering,
National University of Sciences & Technology

Supervisor: Dr. Yasar Ayaz

Committee members: Dr. Mohsin Jamil

Dr. Syed Omer Gilani

Dr. Muhammad Naveed

Outline

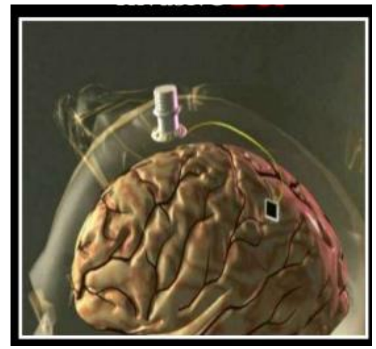
- Introduction
- Contribution
- Classification
- Experimental Results
- Comparison
- Conclusion

Introduction

- Brain Computer Interface (BCI)
  - device that permit brain signals to interact with the environment.
- BCI has been divided into two groups
  - Invasive BCI
    - electrodes are mounted in to the brain skin
  - Non-Invasive BCI
    - electrodes are mounted on the surface
- BCI system used
  - Disabled Persons
  - Non-disable Individuals [2].
- BCI system
  - also used in robotics and biomedical technologies etc. [1], [2].

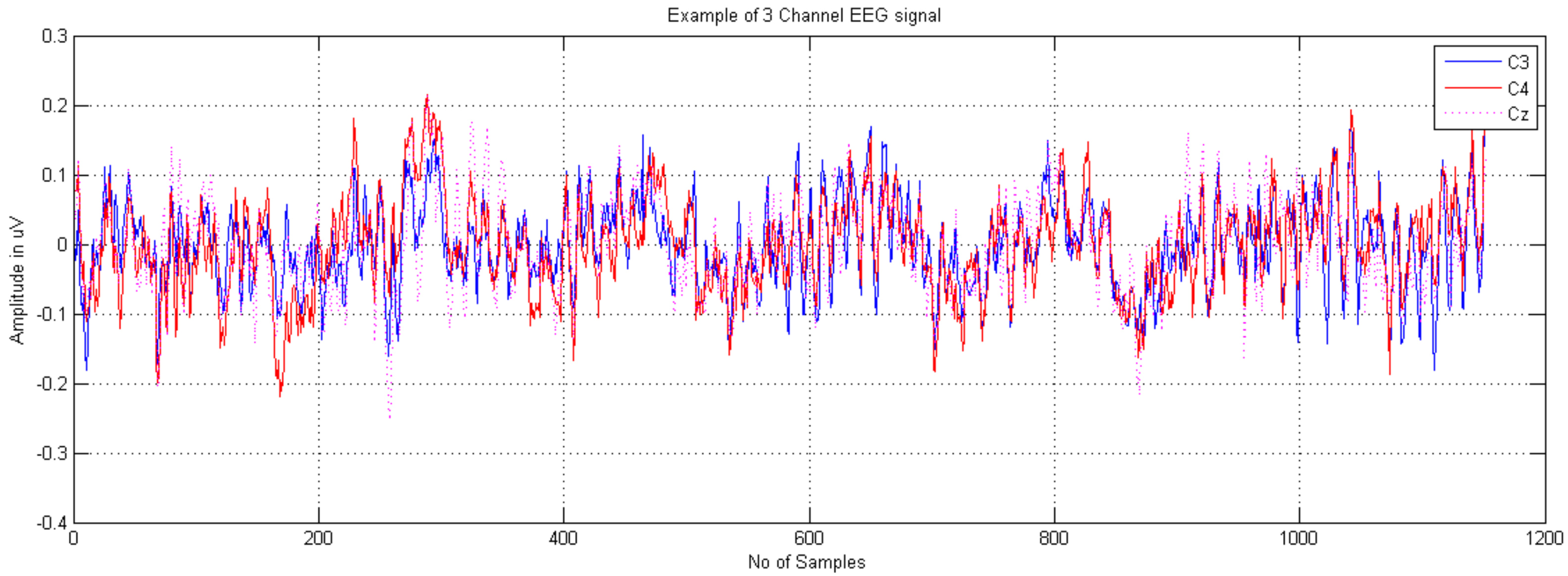
Introduction

- Sources to measure brain activities for BCI.
  - Invasive BCIs
    - electrocorticograpy (ECoG)
    - single micro-electrode (ME)
    - micro-electrode array (MEA)
    - local field potentials (LFPs)
  - Non-invasive BCIs
    - electroencephalography (EEG)
    - magneto encephalography (MEG)
    - Functional Magnetic Resonance Imaging (fMRI)
    - Near Infrared Spectroscopy (NIRS)
- Mostly EEG signal is used
  - the non-invasive EEG electrodes
  - Availability
  - low hardware cost and transferability



Introduction

- Electroencephalogram (EEG)
  - Fluctuations in the voltage caused by the flow of ionic current in the neurons.



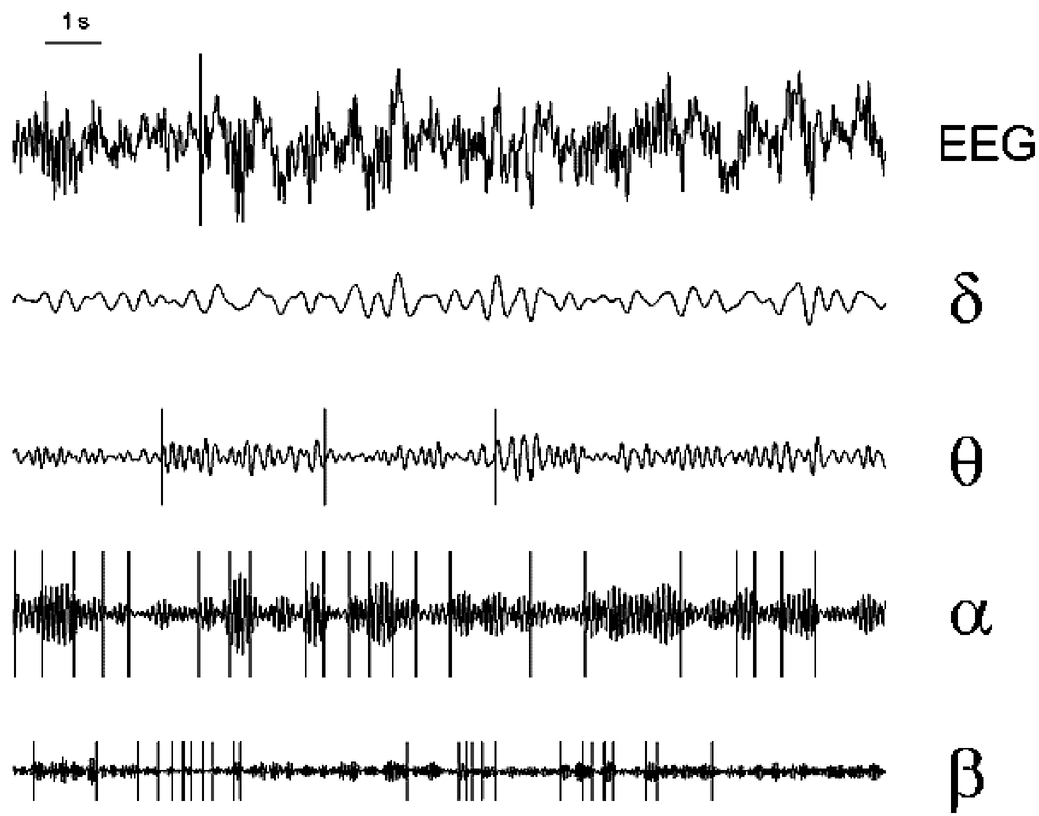
Introduction

- Motor Imagery (MI)
  - EEG signals related to motor movements.
  - External environments [4].
- To convert the MI based EEG signals to BCI input decision and control signals
  1. Signal Enhancement
  2. Feature Extraction
  3. Classification

Introduction

Frequency bands of interest

- Frequency bands
  - δ (0.4 – 4 Hz)
  - θ (4 – 8 Hz)
  - α (8 - 12 Hz)
  - β (12 – 30 Hz)



Images: Courtesy of *“http://brain.bio.msu.ru/papers/chp2000/7.htm”*

Contribution

- Deep learning
  - classify motor imagery based EEG signals
- Signal Enhancement
  - Band-Pass Filter
  - Median filter
- Features Extraction
  - Wavelet Transform (WT)
  - Power Spectral Density
  - Aggregated Signal

Cont.

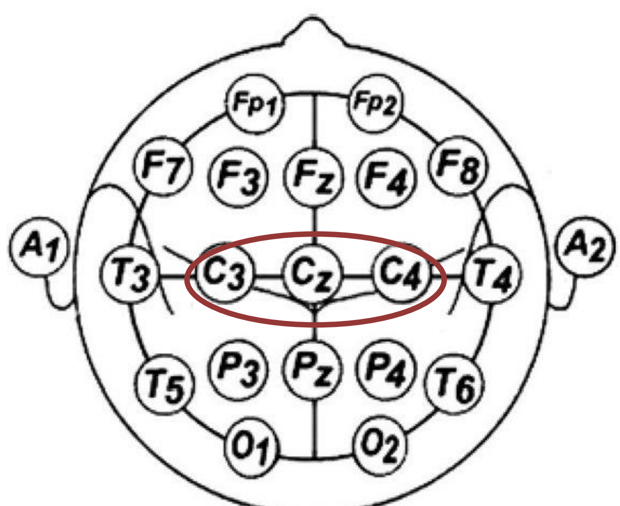
- Classification Results has been compared
  - Support Vector Machines (SVM),
  - k- Nearest Neighbor (k-NN),
  - Linear Discriminant Analysis (LDA),
  - Quadrature Discriminant Analysis (QDA)
  - Self Organizing Maps (SOM) based Neural Networks
  - Deep Belief Networks (Deep Learning)
- Database
  - BCI Competition III named Graz database.

Classification Steps

- Feature Extraction
  - Wavelet Transform (bi-orthogonal 6.8)
  - Power Spectral Density
  - Aggregated Signal
- Feature Reduction
  - Principal Component Analysis
- Classification
  - Support Vector Machine
  - Self Organizing Map
  - Deep Belief Networks

Dataset

- Dataset Details
  - BCI Competition 2003 named Graz data.
  - Collected from a usual subject
    - reaction sitting
      - subject was comforting on chair with supports to its arms.
    - goal is to move a block
      - left and right movement.
  - The electrodes are placed on scalp as on the location as shown



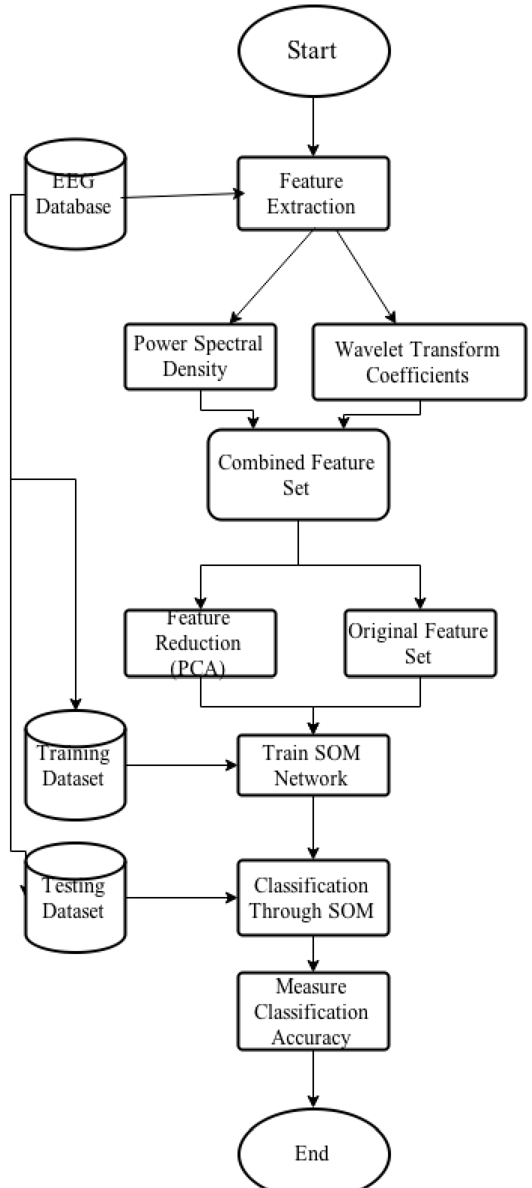
Cont.

- The database contain 280 trails
  - each of 9 seconds
  - Three electrodes Cz, C4 and C3,
  - 140 correspond to training set
  - 140 correspond to testing signals
- The sampling rate is of 128Hz.
- Brain signals are of frequency range of 0.3-40Hz.
- Therefore 0.5-30 Hz band is extracted through a band-pass filter

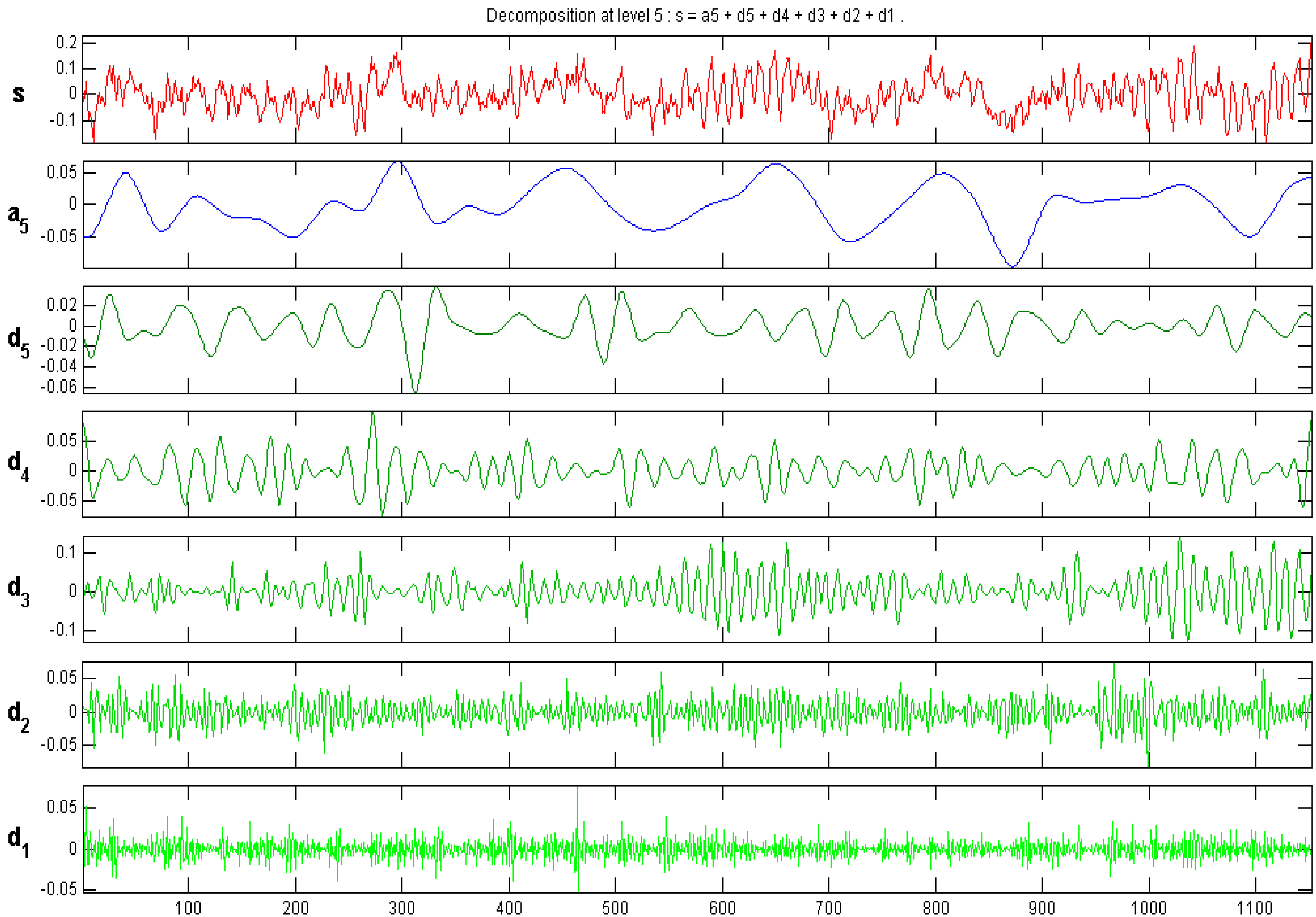
Feature Extraction

- For Self Organizing Map (SOM) based neural network, the best suited features were
  - Wavelet Transform
  - Power Spectral Density
- For Deep Belief Network, the following features has been compared
  - Fourier transform
  - Wavelet transform
  - Aggregated absolute signal

Method for SOM



Wavelet Coefficients



Power Spectral Density

- Power Spectral density (PSD)
  - knowledge of frequency vs. power spreading.
  - autocorrelation of Fourier transform (FT).
- The Welch PSD estimate
  - Hamming window of 64.
- The PSD estimates 8-25 Hz has been extracted [13].
  - 8-12Hz correspond to α band
  - 18-25Hz correspond to the β band.
- Mean power has also been computed for each band.

Feature vector set

- Features extraction is from t = 3s to 9s.
  - Frequency range 0.5-30Hz.
- The feature vector consist of
  - wavelet coefficients
  - PSD estimates for both bands i.e. (8-12Hz and 18-25Hz)
- The data size 871x 140 has been termed as non-reduce feature set.
- Reduce Feature Vector
  - Principal Component Analysis (PCA)
  - Reduced feature Vector 91×140
    - Contains 98 % of the variance

Features	Dimension (Features × samples)
Bior6.8 Wavelet Coefficient	102 ×140
PSD estimate	768 ×140
Mean Power of signal	1 ×140
Total Data Size	871 ×140

Classification

- Final step
  - classify the signal maximum accuracy.
- Different classifiers and compared the results
  - Linear Discriminant Analysis (LDA)
  - Quadratic Discriminant Analysis (QDA)
  - Linear Support Vector Machine (SVM)
  - k-nearest neighbor (kNN)
  - SOM based neural networks
  - Deep Belief Net

Performance Analysis of SOM based Classification

- MATLAB has been used
- Results of both reduced and non-reduced feature vectors has recorded

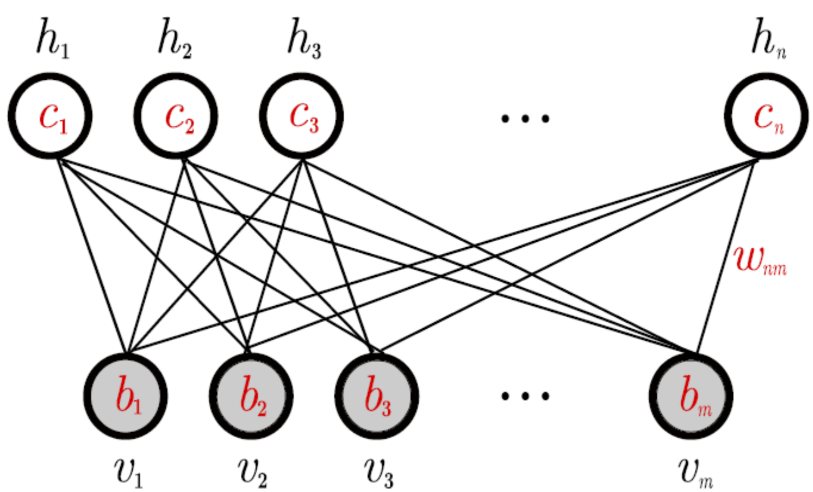
No.	Classification Algorithms	Accuracy (original) %	Accuracy(reduced) %	Reference
1	LDA	80.30	82.64	[13]
2	QDA	80.50	81.70	[13]
3	KNN	77.50	82.90	[12]
4	SVM (Linear)	81.42	81.42	[12]
5	SOM	83.45	84.17	Mine

Deep Learning

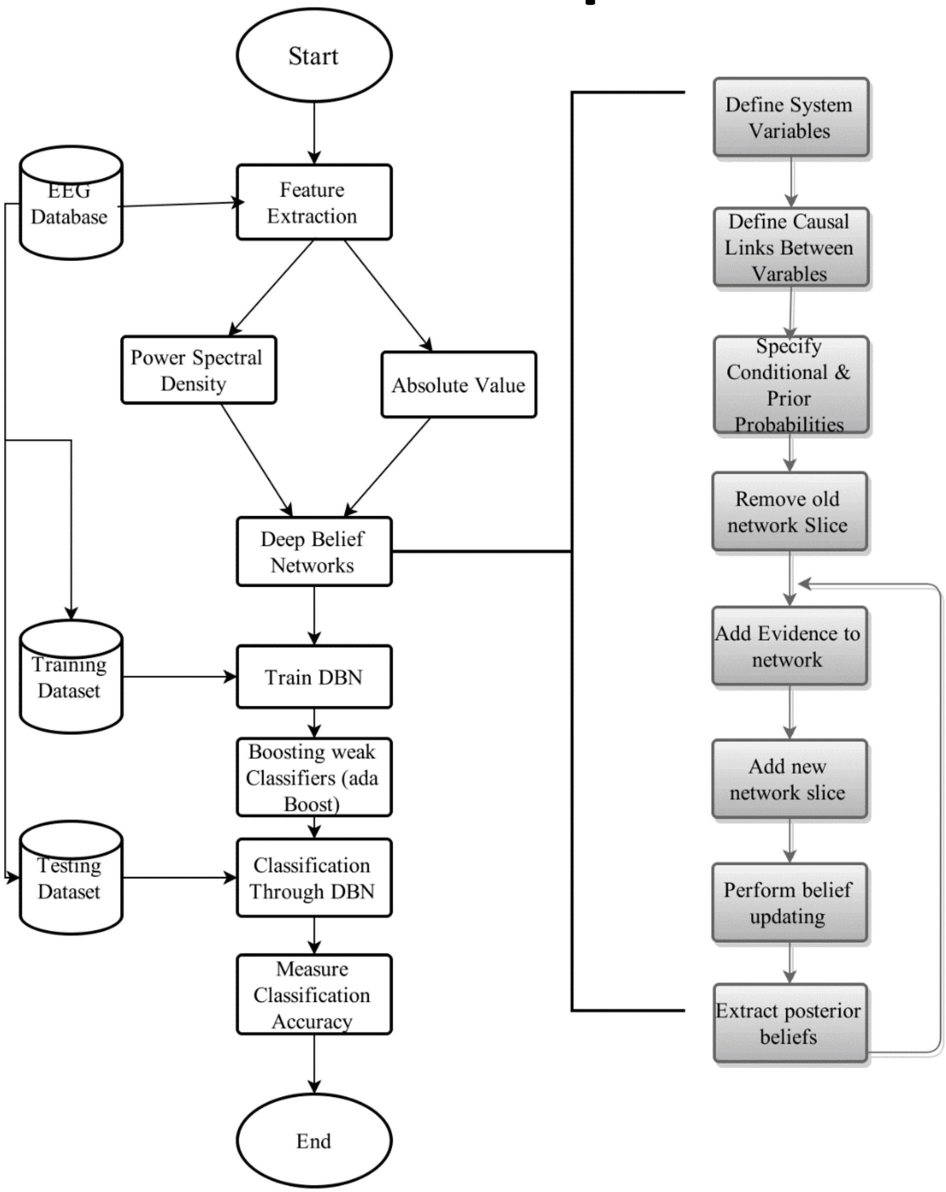
- Deep learning is very lose simulation of brain.
- The algorithms uses several levels and layers of raw data
  - to learn automatically by using a deep structure of neural networks make up of many hidden layers.
- It automatically mine features that are more related to classification
- Involves meaningful information which is not depend upon other features.
- The DBN model is based on a number of Restricted Boltzmann Machine (RBM) which uses unsupervised learning technique [7].

Deep Belief Net (DBN)

- Every single hidden layer in DBN has a Restricted Boltzmann Machine (RBM)
- RBM is trained and weights are assigned to each units.
- The RBM structure is shown in Fig :



Method for Deep Belief Net



Experimental Parameters for DBN

- From 140 total samples 70 samples has been used for training and 70 for testing purpose.
- Each node has been initialized with a random weight.
- The turning parameters were set as:
  - Learning rate for weight and biases = 0.07
  - Momentum=0.5
  - Weight decay = 0.002.
- Range of hidden layers for DBN has been set from 4 to 20 [15]
- The unit size is fixed for every layer

Results

Table 1 Mini Batch Mean Square error using aggregated absolute value

Hidden layers	Epoch						
	1	4	6	7	8	10	Error
4	0.32365	0.22715	0.19225	0.10380	0.05547	0.02137	7%
6	0.49859	0.49618	0.40745	0.26631	0.25521	0.26630	29%
7	0.30189	0.29237	0.20115	0.10559	0.05974	0.01370	8%
9	0.36111	0.2511	0.24729	0.17114	0.14391	0.05030	11%
10	0.38963	0.37796	0.23674	0.21644	0.18168	0.08136	16%
12	0.37125	0.25995	0.19381	0.18716	0.13798	0.07046	11%
14	0.39381	0.27904	0.24590	0.23087	0.21993	0.14084	17%
15	0.35488	0.22000	0.11820	0.11339	0.07837	0.02874	6%
17	0.30835	0.25203	0.21098	0.19075	0.17331	0.07226	11%
18	0.28049	0.25334	0.25358	0.25196	0.23529	0.16987	14%
19	0.27582	0.24876	0.15606	0.14082	0.10631	0.03305	8%
20	0.31849	0.26069	0.24548	0.24479	0.25184	0.21088	16%
Epoch wise	30%	23%	17%	13%	10%	4%	7%

Results

Table 2 Full Batch Mean square Error using Aggregated Absolute Value with different number of hidden layers

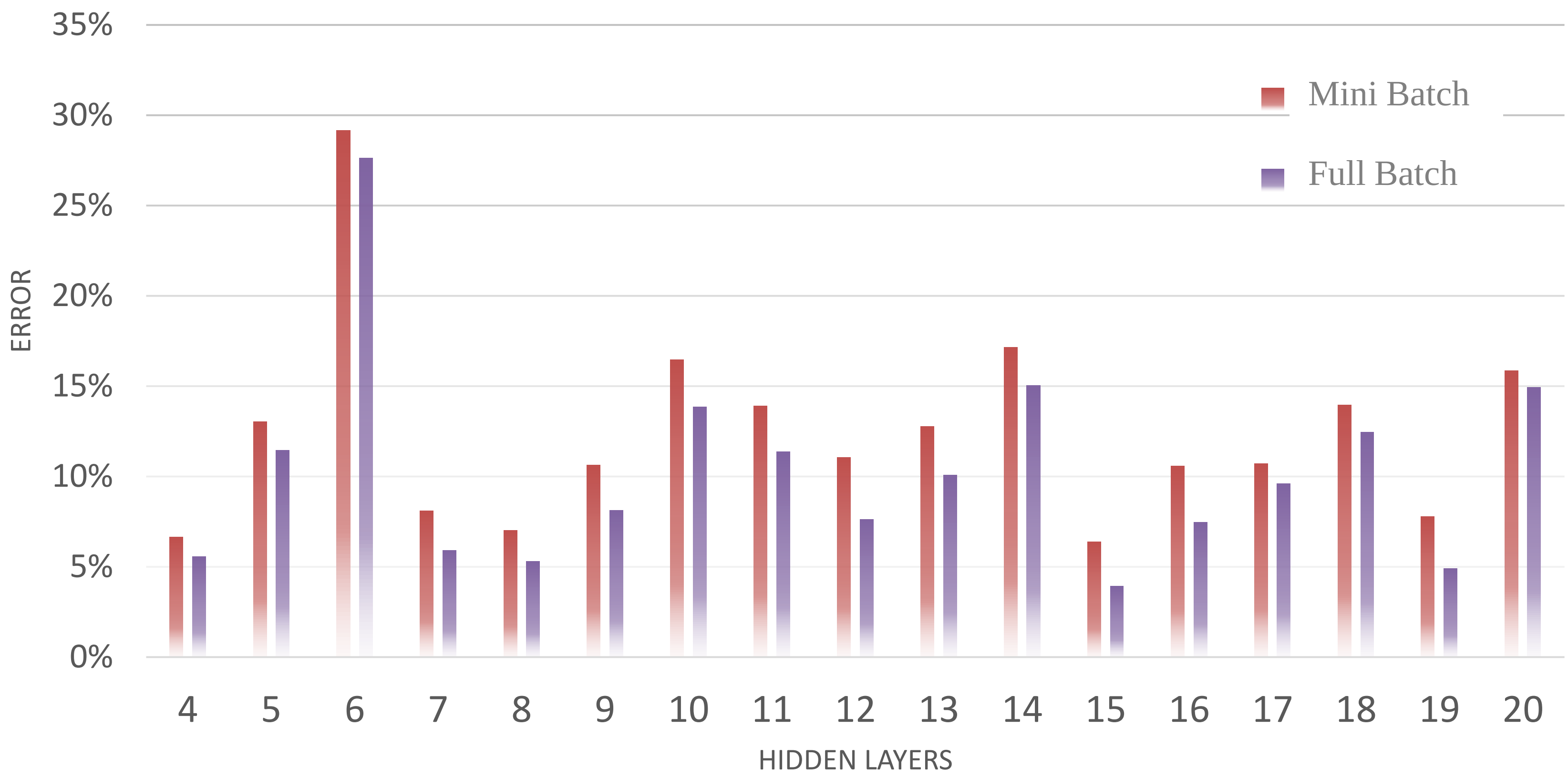
Hidden layers	Epoch						
	1	4	6	7	8	10	Error
4	0.328874	0.186094	0.127025	0.055325	0.032907	0.010847	6%
5	0.303948	0.250407	0.211553	0.251637	0.123172	0.040178	11%
6	0.498398	0.494803	0.264515	0.256402	0.254118	0.222160	28%
7	0.260469	0.229944	0.100652	0.055693	0.041782	0.012354	6%
9	0.279942	0.233898	0.164947	0.198849	0.100792	0.025755	8%
10	0.375353	0.353728	0.192056	0.174571	0.112999	0.050374	14%
12	0.249831	0.244736	0.175301	0.140486	0.101412	0.045612	8%
14	0.399336	0.280878	0.238957	0.242188	0.224081	0.114004	15%
15	0.238821	0.204236	0.087956	0.130813	0.035063	0.035500	4%
16	0.412291	0.238595	0.162861	0.115279	0.067042	0.023590	7%
18	0.249528	0.253878	0.240352	0.221536	0.210868	0.127481	12%
19	0.285597	0.206465	0.125359	0.103638	0.066685	0.019502	5%
20	0.253958	0.273584	0.243070	0.248365	0.228325	0.184926	15%
Epoch wise	27%	21%	13%	12%	7%	2%	5%

Results

Table 3 Full Batch Classification Error Using PSD

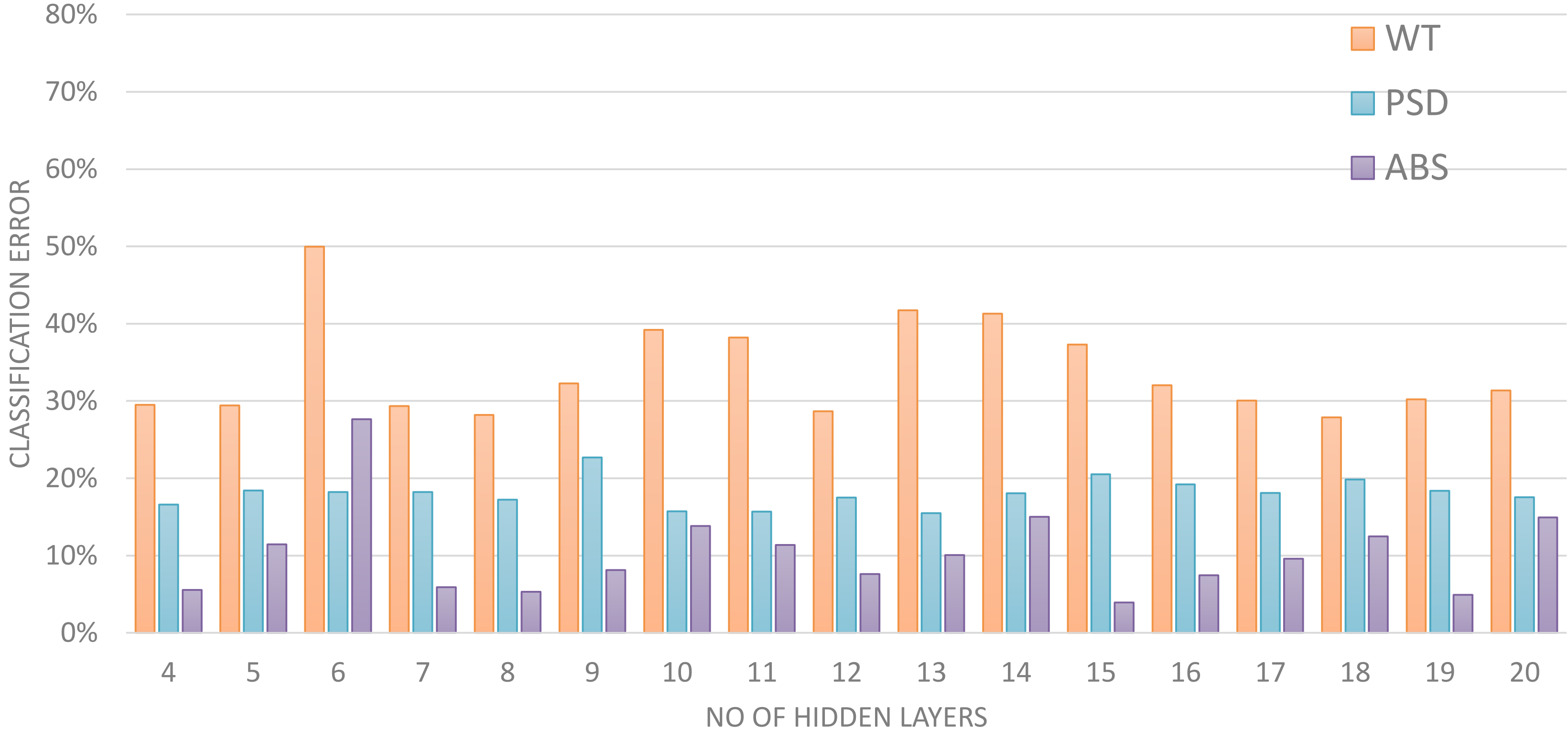
Hidden Layers	Epoch						
	1	3	5	6	8	10	Error
4	0.158124	0.197538	0.171107	0.151335	0.150454	0.155412	17%
5	0.276637	0.161026	0.152534	0.166176	0.18365	0.149961	18%
8	0.151637	0.157858	0.16407	0.159322	0.194539	0.151622	17%
9	0.299362	0.275149	0.286839	0.152936	0.151099	0.161565	23%
10	0.156181	0.150529	0.153986	0.150587	0.15969	0.153361	16%
13	0.150986	0.150162	0.156962	0.178769	0.151577	0.150494	15%
14	0.297748	0.189421	0.150261	0.169156	0.150458	0.153936	18%
15	0.396984	0.165612	0.158541	0.183391	0.156187	0.15078	21%
16	0.301985	0.229711	0.168155	0.155346	0.162087	0.155824	19%
17	0.240276	0.201087	0.173229	0.161217	0.204058	0.159594	18%
18	0.386699	0.167499	0.162614	0.153515	0.179621	0.159482	20%
20	0.239277	0.161283	0.160649	0.156961	0.156937	0.217208	18%
Epoch Wise	25%	18%	18%	16%	17%	16%	18%

Comparison of Classification



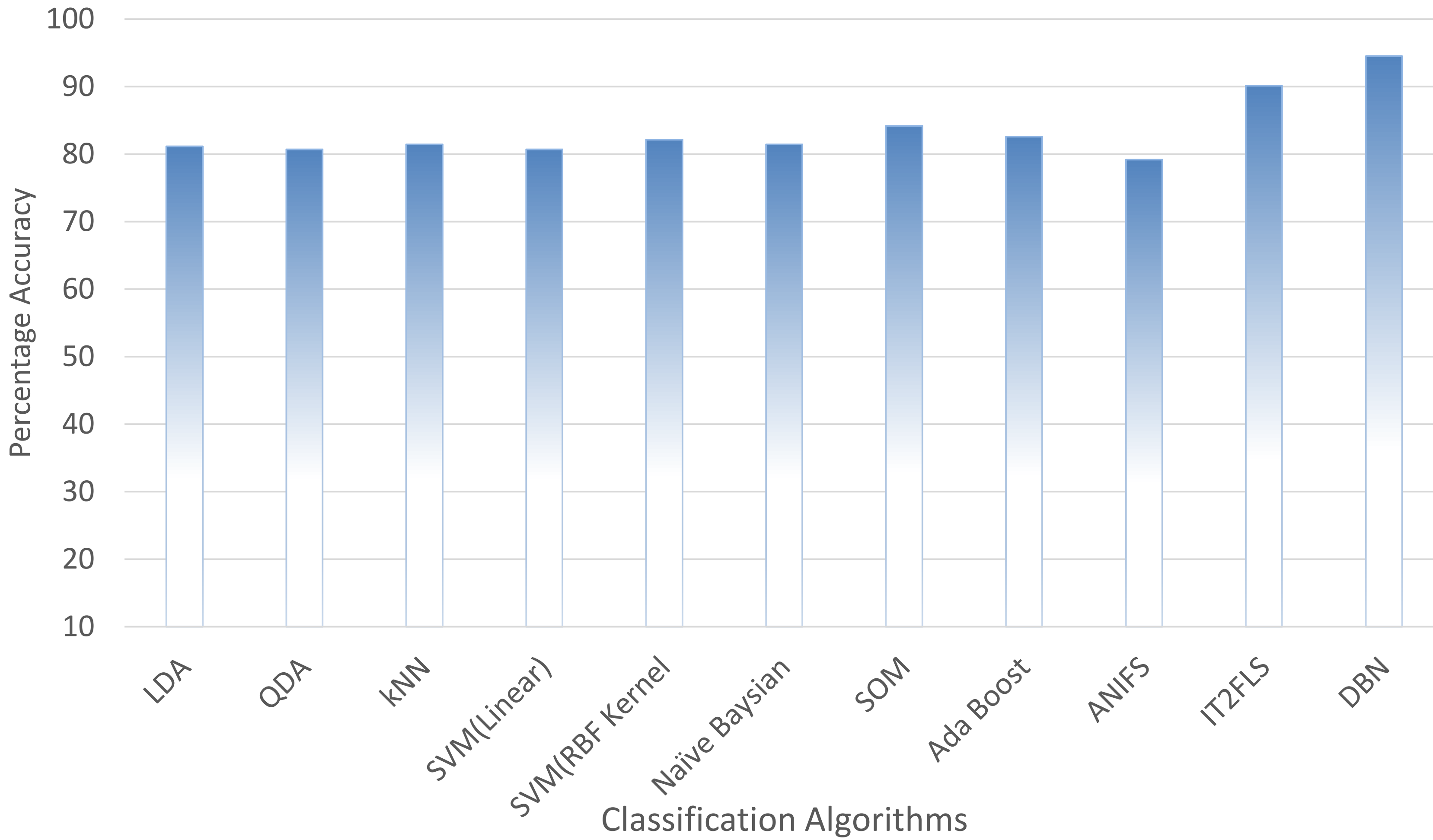
Comparison of mini and full batch classification error Based on Hidden unit

Comparison of three different features set



Classification error of three different features set

Overall Comparison



Comparison of different classification Algorithms [14]

Conclusion

- In this thesis
  - An efficient and relatively new classification technique
- Two Learning Techniques has been implemented
  - SOM based Neural Networks
  - Deep Belief Nets (DBN)
- For SOM
  - The best features set include Bior 6.8 Wavelet transform, PSD approximation and mean power.
  - SOM gave the highest classification efficiency compared to LDA, QDA and SVM.
  - reduced feature set by applying PCA has increased as compare to the non-reduced feature sets.
  - The bi-orthognal wavelet transform has also increased the classification accuracy.

Conclusion

- For DBN
  - The result are very impressive
  - The average minimum classification errors are
    - 5% in Full Batch Classification
    - 8% in Mini Batch Classification
  - The result of Deep Learning Algorithms are quite well as compare to others
  - The main problem with deep learning is complexity and difficult to optimize on stand alone platforms.

Future Direction

- Our future plan
  - Design a system that has the ability to online classify motor imagery EEG signals and able to control a mobile robot in a real environment.
  - Development of optimized deep learning algorithm for low end devices.
  - Develop an algorithm that has the capability of Adaptive feature selection

Research Publication

- Baig, Muhammad Zeeshan, Yasar Ayaz, Waseem Afzal, Syed Omer Gillani, Muhammad Naveed, and Mohsin Jamil. “A Comparative Analysis of Motor Imagery based EEG Signals Classification Algorithms with Deep Belief Networks.” *In BioMed Research International Journal (**Submitted**) (ISI Indexed)*
- Baig, Muhammad Zeeshan, Yasar Ayaz, Waseem Afzal, Syed Omer Gillani, Muhammad Naveed, and Mohsin Jamil. “Motor Imagery Based EEG Signal Classification Using Self Organizing Maps.” In Science International Journal (**Published**) (ISI Indexed)
- Baig, Muhammad Zeeshan, Ehtasham Javed, Yasar Ayaz, Waseem Afzal, Syed Omer Gillani, Muhammad Naveed, and Mohsin Jamil. "Classification of left/right hand movement from EEG signal by intelligent algorithms." In *Computer Applications and Industrial Electronics (ISCAIE), 2014 IEEE Symposium on*, pp. 163-168. IEEE, 2014. (**Published**)

Reference

1. R. Sitaram, A. Caria, R. Veit, T. Gaber, G. Rota, A. Kuebler, and N. Birbaumer, “FMRI brain-computer interface: A tool for neuroscientific research and treatment”, Computational intelligence and neuroscience, Jan. 2007.
2. S.M. Coyle, T.E. Ward, and C.M. Markham, “Brain-computer interface using a simplified functional near-infrared spectroscopy system”, Journal of neural engineering, vol. 4, Sep. 2007, pp. 219-226.
3. N.J. Hill, T.N. Lal, M. Schr, T. Hinterberger, G. Widman, C.E. Elger, B. Sch, and N. Birbaumer, “Classifying Event-Related Desynchronization in EEG ,ECoG and MEG signals”, Interface, 2006, pp. 404-413
4. J. a Pineda, “The functional significance of mu rhythms: translating "seeing" and "hearing" into "doing"”, Brain research reviews, vol. 50, 2005, pp. 57-68.
5. Gorge, H.S.O.: A fast learning algorithm for deep belief nets. Neural Computation, 1527–1554 (2006)
6. Wu, S., Wu, W.: Common Spatial Pattern and Linear Discriminant Analysis for Motor Imagery Classification. In: 2013 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB), pp. 146–151 (2013)
7. Yoshua, B., Pascal, L.: Greedy layer-wise training of deep networks. NIPS (2006)
8. Yohei, Y., Mitsukura, Y.: Hemodynamic characteristics for improvement of EEG-BCI performance. In: 2013 The 6th International Conference on Human System Interaction (HSI), pp. 495–500 (2013)
9. Wolpaw, J., Birbaumer, N.: Brian-computer interfaces for communication and control. Clinical Neurophysiology, 767–791 (2002)
10. Yohei, T., Yasue, M.: Boosted Network Classifiers for Local Feature Selection. IEEE Transactions on Neural Networks and Learning Systems, 1767–1778 (2012)
11. Kirkup, L., Searle, A.: EEG-based system for rapid on-off switching without prior learning. Medical and Biological Engineering and Computing, 504–509 (2007)
12. Hochberg, L., Serruya, M.: Neuronal ensemble control of prosthetic devices by a human with tetraplegia. Nature, 164–170 (2006)
13. Bhattacharyya, S., Khasnobish, A., Konar, A., Tibarewala, D. N., & Nagar, A. K. (2011, April). Performance analysis of left/right hand movement classification from EEG signal by intelligent algorithms. In Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB), 2011 IEEE Symposium on (pp. 1-8). IEEE.
14. Nguyen, T., Khosravi, A., Creighton, D., & Nahavandi, S. (2015). EEG signal classification for BCI applications by wavelets and interval type-2 fuzzy logic systems. Expert Systems with Applications, 42(9), 4370-4380.

15. An, X., Kuang, D., Guo, X., Zhao, Y., & He, L. (2014). A Deep Learning Method for Classification of EEG Data Based on