# Sleep Stages Classification from Electroencephalographic Signals based on Unsupervised Feature Space Clustering

Iosif Mporas, Anastasia Efstathiou, Vasileios Megalooikonomou

Multidimensional Data Analysis and Knowledge Management Laboratory,
Dept. of Computer Engineering and Informatics, University of Patras
26500 Rion-Patras, Greece
imporas@upatras.gr

**Abstract.** In this article we present a methodology for the automatic classification of sleep stages. The methodology relies on short-time analysis with time and frequency domain features followed by unsupervised feature subspace clustering. For each cluster of the feature space a different classification setup is adopted thus fine-tuning the classification algorithm to the specifics of the corresponding feature subspace area. The experimental results showed that the proposed methodology achieved a sleep stage classification accuracy equal to 92.53%, which corresponds to an improvement of approximately 3% compared to the best performing single classifier without applying clustering of the feature space.

**Keywords:** sleep stages, electroencephalography, clustering.

## 1 Introduction

The modern lifestyle at work as well as at off-work daily activities, especially in the case of developed countries, is frequently characterized as stressful, which can result in mental distress. Population suffering from sleep problems is increasing and sleep disordering becomes one of the popular health problems. This affects the physical and mental health of people, in contrast to effective sleep which results to comfort and relief from stress. The quality of sleep is more important than the quantity and the falling asleep period is important in obtaining good quality sleep [1]. The inability either to sleep or to stay awake is a significant health problem which may result in an abnormal everyday living.

In order to diagnose and offer appropriate therapy to sleep disorders, the understanding of the brain and body mechanism during sleep is necessary. Sleep is a losing state as temporary, partial, and periodic in the form of that can be returned with various forced stimulus of the communication of the organism with the environment [2]. It can also be defined by the decreasing of motor activity, the decreasing of response with stimulus, and to be easy recycling as behavioral [3, 4].

Neuroscientists have described the human sleep, more than one century ago after studies using electroencephalography (EEG), as a succession of five recurring stages.

These stages are the rapid eye movement stage (REM) and the non-rapid eye movement stages (NREM), which are distinguished in N1, N2, N3 and N4 NREM stages. The awake status is considered as a separate stage and is not included in the above five ones. These stages are characterized by rapid changes in the amplitude and the frequencies (rhythms) of the electroencephalographic (EEG) signal [5]. Specifically, N1 presents a frequency transition of the brain from alpha waves (approx. 8–13 Hz, common in the awake state) to theta waves (approx. 4-7 Hz). N2 is characterized by sleep spindles (approx. 11-16 Hz) and K-complexes [6]. In N3 stage a minimum of 20% delta waves (approx. 0.5-2 Hz) appear, having a peak-to-peak amplitude of more than 75 μV, while in N4 stage delta waves reach 50%. REM stage accounts for 20–25% of total sleep time duration in most human adults and presents a rapid low-voltage EEG (higher frequency saw-tooth waves). The study and the analysis of the sleep stages from sleep scientists is essential for the treatment and therapy of sleep related disorders. It typically results in a hypnogram [23], i.e. a form of polysomnography (PSG) that graphically represents the stages of sleep as a function of time.

The analysis and evaluation of sleep is performed by expert neurologists using EEG signals. Sleep recordings usually last approximately 6-8 hours and thus visual investigation of the EEG signals manually by a sleep scientist is quite tedious and time-consuming. Moreover, manual investigation of the EEG sleep recordings may result to biased/subjective analysis conclusions since it will rely on the degree of experience of the sleep expert. Additionally, the need for sleep experts sets manual analysis of sleep recordings as an expensive solution. Due to the above mentioned disadvantages of manual analysis of sleep recordings, computer-based solutions for the automatic sleep stage scoring have been proposed based on the recent advances in information technology and specifically in the areas of statistical signal modeling, pattern recognition and machine learning.

Most methodologies found in the literature are based on short-time decomposition of the EEG signals to frames (or epochs). The epochs are represented by spectral analysis methods [2, 12, 14, 15], wavelet transformation [7, 8], autoregressive analysis [5, 11], band-specific energy [9, 10], entropy [10, 13], etc. Except from the EEG signal, several studies report analysis of sleep stages using electrocardiogram (ECG) [1, 9, 16] signals. Furthermore, in [10, 15] the study of sleep and wakefulness was performed using polysomnographic (PSG) data, i.e., concurrent recordings of multiple sensors including EEG, ECG, electro-oculogram (EOG) and electromyogram (EMG). For the automatic scoring of sleep stages each epoch is labeled with one of the sleep stage classes by classifying the corresponding feature vectors. Several well-known machine learning algorithms for classification have been used in this task, such as the support vector machines (SVMs) [9, 13], the artificial neural networks (NNs) [7, 8, 10, 12, 16], the decision trees (DTs) [2] and other classification methods such as the Gaussian mixture models (GMMs) [14], the hidden Markov models (HMMs) [17], the Kullback-Leibler divergence (KL) [5] the Bayesian classifier (BN) and the k-nearest neighbors (knn) [10].

In this article we present a methodology for automatic classification scoring of the sleep stages from electroencephalographic data, based on feature subspace clustering and use of cluster-specific classifiers for the sleep stage scoring. The remaining of this article is organized as follows: In Section 2 we explain the proposed methodology for automatic sleep stage classification. In Section 3 we offer a description of the data used in the evaluation and describe the experimental setup. In Section 4 we describe the experimental results and conclude in Section 5.

## 2 Sleep Stages Classification using Feature Space Clustering

The proposed methodology for sleep stage classification is based on the use of time and frequency based EEG signal features in order to exploit the differences in the frequencies and the amplitudes of the EEG signal among the sleep stages, as described in Section 1. In particular, during the training phase of the methodology we perform short-time analysis of the EEG signal using widely used time-domain and frequency-domain features and perform unsupervised clustering of the feature vector space. Moreover for a number of dissimilar classifiers we examine their discriminative ability per cluster and assign the best-performing classifier per cluster. In the test phase, each EEG epoch is represented by the time-domain and frequency-domain features used in training and is classified to one of these feature space clusters as they were modeled during training. Subsequently, the corresponding cluster-specific classifier is used to assign a sleep stage score to the corresponding EEG epoch. The concept of the proposed methodology is illustrated in Figure 1.

As can be seen in Figure 1, during the training phase a bootstrap set of EEG data with known hypnogram labeling, i.e., known time intervals per sleep stage is used. The training EEG signal is initially preprocessed with frame blocking of constant length $w$ samples and time-shift $t \leq w$ samples, thus decomposing the EEG signal

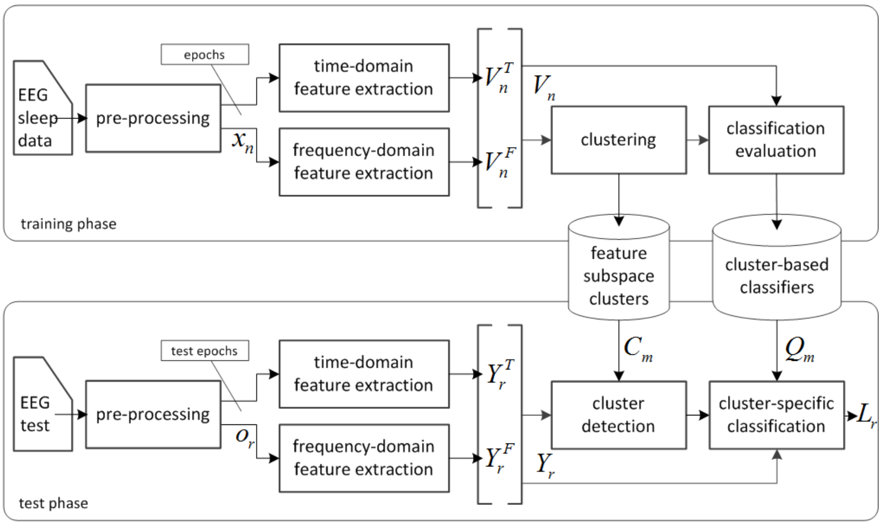


**Fig. 1.** Block diagram of the proposed methodology for sleep stage classification from EEG data with unsupervised feature space clustering.

to a sequence of $N$ epochs $X = \{x_n\}$, $x_n \in \mathbb{R}^{w}$, with $1 \le n \le N$. Here we consider single channel EEG analysis. Each epoch is processed by time domain parameterization algorithms producing a corresponding feature vector, $V_n^T \in \mathbb{R}^{|V_n^T|}$, $1 \le n \le N$, and by frequency domain parameterization algorithms producing a corresponding feature vector, $V_n^F \in \mathbb{R}^{|V_n^F|}$, $1 \le n \le N$. The time and frequency domain feature vectors of each epoch are concatenated into a common feature vector, i.e. $V_n^T \| V_n^F = V_n \in \mathbb{R}^{|V_n^T|+|V_n^F|}$, $1 \le n \le N$. Subsequently, unsupervised clustering of the feature vector space to $M$ subspace clusters, $C_m$, $1 \le m \le M$, is applied using a clustering algorithm. For each of the $M$ subsets of epochs belonging to the $m$-th cluster, we evaluate $K$ dissimilar classification algorithms, $S_k$, $1 \le k \le K$, in order to extract the best performing, $Q_m$, classifier per cluster, i.e.

$$Q_m = \arg\max_k \left(C_m, S_k\right) \qquad (1)$$

where $Q_m \in \{S_k\}$.

During the test phase a test EEG sleep signal recording, $O$, with unknown sleep stage annotation is preprocessed similarly to the training phase. The corresponding epochs, $O = \{o_r\}$, $1 \le r \le R$, $o_r \in \mathbb{R}^{w}$, are processed by time-domain and frequency-domain feature extraction algorithms, identical to the training phase, producing the feature vectors $Y_r^T$ and $Y_r^F$, respectively. The concatenated test feature vectors $Y_r^T \| Y_r^F = Y_r \in \mathbb{R}^{|V_n|}$ are subsequently processed by a cluster detector, which classifies each test vector, $Y_r$, to one of the $M$ clusters, as they were modeled from the unsupervised clustering of the training data. After assigning a cluster label to each of the test epochs, i.e. $Y_r \in C_m$, $1 \le r \le R$, the corresponding cluster-specific sleep stage classification algorithm, $Q_m$, is used to score each of the test epochs with a sleep stage label, i.e.

$$L_r = f\left(Y_r, Q_m\right),\ Y_r \in C_m \qquad (2)$$

where $L_r$ is the sleep stage label of the $r$-th test epoch.

The above architecture for sleep stage scoring can be extended to more channels without any loss of generality. The use of unsupervised clustering of the feature space may support the used classification engines in fine-tuning their free parameters for sleep stages separation in a subspace of the features with less varying characteristics.

## 3 Experimental Setup

The sleep stage classification architecture presented in Section 2 was evaluated using "The Sleep EDF Database" [18] provided by Physionet [19]. It is a collection of 61 polysomnograms (PSGs) with accompanying hypnograms. The recordings were obtained from Caucasian males and females (21 - 35 years old) without any medication and they contain bipolar Fpz-Cz and Pz-Oz EEG channels according to the 10-20 system [24], sampled at 100 Hz. The labeling of the electrodes indicate the position of the EEG electrodes: the letters F, T, C, P and O stand for frontal, temporal, central, parietal, and occipital lobes, respectively, a "z" (zero) refers to an electrode placed on the midline, even numbers (2,4,6,8) refer to electrode positions on the right hemisphere, whereas odd numbers (1,3,5,7) refer to those on the left hemisphere. The Fpz-Cz channel was used here. Hypnograms were manually scored according to Rechtschaffen & Kales based on Fpz-Cz/Pz-Oz EEG channels.

During pre-processing the EEG recordings were frame blocked to epochs of 15, 30, 45 and 60 seconds length, without time-overlap between successive epochs. For each epoch well-known and widely used time and frequency domain features were extracted [20]. In particular, the EEG waveform was parameterized using the following features: (i) time-domain features: minimum value, maximum value, mean, variance, standard deviation, percentiles (25%, 50%-median and 75%), interquartile range, mean absolute deviation, range, skewness, kyrtosis, energy, Shannon's entropy, logarithmic energy entropy, number of positive and negative peaks, zero-crossing rate, and (ii) frequency-domain features: 6-th order autoregressive-filter (AR) coefficients, power spectral density, frequency with maximum and minimum amplitude, spectral entropy, delta-theta-alpha-beta-gamma band energy, discrete wavelet transform coefficients with mother wavelet function Daubechies 16 and decomposition level equal to 8, thus resulting to a feature vector of dimensionality equal to 55.

For the clustering step we used the expectation-maximization (EM) algorithm. Both for the training and the test phase classification we relied on dissimilar algorithms in order to maximize complementarity. Specifically, we used the k-nearest neighbors classifier with linear search of the nearest neighbor and without weighting of the distance also known as instance based classifier (IBk), the pruned C4.5 decision tree (C4.5) with 3 folds for pruning and 7 for growing the tree, the multilayer perceptron neural network (MLP) with an architecture of 3-layers, and the support vector machine (SVM) using the sequential minimal optimization algorithm and radial basis function kernel. The values of the parameters $C$ and $\gamma$ for the kernel were empirically set to 15 and to 0.1 respectively after grid search. For the implementation of the algorithms we relied on the WEKA toolkit implementations [21].

## 4 Experimental Results

The sleep stage classification methodology using unsupervised feature space clustering presented in Section 2 was evaluated following the experimental setup described in Section 3. In order to evaluate the performance of the proposed methodology we examined the percentage of correctly classified epochs. We followed a ten-fold cross validation setup, in order to avoid overlap between training and test datasets.

As a first step, we examined the performance of the four evaluated classification algorithms with different size, $w$, of frame length (epoch). The results (in percentages) are tabulated in Table 1. The best performing setup is shown in bold. In all evaluated epoch sizes there is no overlap between successive epochs.

**Table 1.** Sleep stage classification accuracy for different classification algorithms using different epoch lenghts.

| Classifier | $w$=15s | $w$=30s | $w$=45s | $w$=60s |
|---|---|---|---|---|
| IBk | 85.94 | 87.50 | 87.10 | 86.32 |
| C4.5 | 84.13 | 86.31 | 85.92 | 84.73 |
| SVM | 87.66 | **89.44** | 88.84 | 88.31 |
| MLP | 87.12 | 88.02 | 87.63 | 87.09 |

As can be seen in Table 1, the best performing algorithm is the SVM with best achieving accuracy equal to 89.44%. The best performing SVM is followed by the second discriminative algorithm, MLP, with an accuracy of 88.02%. The other two algorithms achieved worse performance. With respect to the size to the epoch, the 30 seconds length achieved the best performance across all evaluated classification algorithms. The experimental results show that the use of longer epochs slightly decreases the performance, while the use of shorter epochs significantly increases the error rate. The superiority of SVMs is probably owed to the fact that these classifiers circumvent the "curse of dimensionality" and the fact that SVM training always finds a global solution [22], in contrast to other classifiers (e.g. neural networks) where many local minima usually exist and thus are sensitive to the distributional specifics of the training data. More than 65% of the misclassified epochs were found close to the transitions of the sleep stages and mainly at the beginnings of each sleep stage. The most misclassified is the sleep stage N1 followed by N3, which is in agreement with [2, 7, 8, 12]. Although due to different data and experimental setup no direct comparison to the reported in the literature sleep stage classification performance (varying approximately from 80% to 95% [2, 5, 7, 8, 10]) can be done, the above reported baseline performance is comparable to the literature.

In a second step we examined the performance of the sleep stage classification methodology with unsupervised feature clustering. Specifically, we applied the EM clustering algorithm for 2 up to 4 clusters. The performance of the cluster-based methodology was evaluated using the best performing 30 seconds epoch size. The experimental results are tabulated in Table 2. The best sleep stage classification performance is indicated in bold. The results of Table 1, which correspond to the case of M=1 cluster (i.e. without clustering) are duplicated in Table 2. The last row corresponds to the results for sleep stage classification using the best per cluster classifier.

**Table 2.** Sleep stage classification accuracy for different classification algorithms using different number M of feature space clusters.

| Classifier | M=1 | M=2 | M=3 | M=4 |
|---|---|---|---|---|
| IBk | 87.50 | 89.33 | 90.58 | 89.96 |
| C4.5 | 86.31 | 88.79 | 89.47 | 89.04 |
| SVM | 89.44 | 91.56 | 92.39 | 91.30 |
| MLP | 88.02 | 91.31 | 92.26 | 91.28 |
| Best Cluster-based | - | 91.60 | **92.53** | 91.63 |

The experimental results in Table 2 show a significant improvement of the sleep classification accuracy when using feature subspace clusters compared to the results of single classification algorithms without clustering shown in the first column. In detail, the best performance was achieved when clustering the feature space to 3 subspaces. The accuracy in this case was equal to 92.53%, which results to an improvement of 3.09% in terms of absolute performance. The increase of the number of clusters was not followed by an increase in the sleep stage classification accuracy. To some degree, this is probably owed to the amount of data, since by increasing the number of clusters the available data for training the cluster-based classification models are decreasing. Moreover, the use of many clusters increases the test feature vectors that are incorrectly assigned to clusters. Except for the best per cluster performing algorithm (in the last row of Table 2), the use of feature space clustering increased the sleep stage classification performance for all evaluated algorithms. Specifically, the SVM classifier, which was the best performing algorithm when no clustering was applied, improved its performance by approximately 3%. The MLP algorithm was improved by approximately 4% and both the IBk and C4.5 algorithms by approximately 3%. In all evaluated cases the use of feature space clustering resulted in less misclassifications around the sleep stage transitions, while the improvement comparing to the baseline setup was not found significant for the middle of each sleep stage. This is probably owed to the fact that the application of feature space clustering restricts the classification to a less varying subspace area, which is more intense close to the sleep stage transitions.

The evaluation showed that in all cases the use of feature space clustering before training the sleep stage classification algorithms results in more robust models. When the training data are restricted to a subarea of the feature space the free parameters of the classifier are better fine-tuned to the diversity among the sleep stages and not to other potential variations of the feature vectors. This results to classification models with better discriminative ability among the EEG sleep stages.

## 5 Conclusion

We presented a methodology for the robust classification of sleep stages from EEG signals. The methodology exploits time-domain and frequency-domain features and trains sleep stage classification models on subareas of the feature space, after unsupervised clustering of the training data. The experimental results showed the validity of the proposed methodology, since the accuracy was improved by more than 3% when using cluster-specific sleep stage classification models instead of models trained with data from the whole feature vector space.

The use of clusters of the feature space offers advantage to the classification algorithms to fine-tune their free parameters to the specific characteristics of each cluster, i.e. to space areas which typically will have lower variation and will be less sparse. Finally, the specific and dissimilar characteristics of the feature data within each cluster can better be modeled by different classification algorithms.

## 6 Acknowledgment

The research reported in the present paper was partially supported by the ARMOR Project (FP7-ICT-2011-5.1-287720) “Advanced multi-paRametric Monitoring and analysis for diagnosis and Optimal management of epilepsy and Related brain disorders”, co-funded by the EC under the 7th Framework Programme and the BioMedMine Project "Mining Biomedical Data and Images: Development of Algorithms and Applications" funded through the Operational Program "Education and Lifelong Learning" of the NSRF-Research Funding Program: Thales. Investing in knowledge society through the European Social Fund.

## 7 References

1. H. Hagiwara, "Estimation of sleep stage in the falling asleep period using a Lorenz plot of ECG RR intervals", In Proc. of the 31st Annual International Conference of the IEEE EMBS, 2009, pp. 2510-2513.
2. S. Gunes, K. Polat, S. Yosunkaya, "Efficient sleep stage recognition system based on EEG signal using k-means clustering based feature weighting", Expert Systems with Applications, vol. 37, 2010, pp. 7922–7928.
3. K. Polat, S. Yosunkaya, S. Gunes, "Comparison of different classifier algorithms on the automated detection of obstructive sleep apnea syndrome", J. of Medical Systems, vol. 32(3), 2008, pp. 243–250.
4. A. Rechtschaffen, A. Kales, "A manual of standardized terminology, techniques and scoring system for sleep stages of human subject", Washington, DC: US Government Printing Office, National Institute of Health Publication, 1968.
5. I. Zhovna, I.D. Shallom, "Automatic detection and classification of sleep stages by multichannel EEG signal modeling", In Proc. of the 30th IEEE EMBS Conference, 2008.
6. I. Mporas, P. Korvesis, E.I. Zacharaki, V. Megalooikonomou, "Sleep spindle detection in EEG signals combining HMMs and SVMs", Artificial Intelligence: Methods and Applications Lecture Notes in Computer Science, vol. 8445, 2014, pp. 442-447.
7. C.-C. Chiu, B. Huy Hai, S.-J. Yeh, "Sleep stages recognition based on combined artificial neural network and fuzzy system using wavelet transform features", In proc. of the 4th Int. Conf. on Biomedical Engineering, IFMBE Proceedings, vol. 40, 2013, pp. 72–76.
8. F. Ebrahimi, M. Mikaeili, E. Estrada, H. Nazeran, "Automatic sleep stage classification based on EEG signals by using neural networks and wavelet packet coefficients", In Proc. of the 30th IEEE EMBS Conference, 2008.

9. S. Yu, Xi Chen, B. Wang, X. Wang, "Automatic sleep stage classification based on ECG and EEG features for day time short nap evaluation", In Proc. of the 10th World Congress on Intelligent Control and Automation, 2012.
10. L. Zoubek, S. Charbonnier, S. Lesecq, A. Buguet, F. Chapotot, "Feature selection for sleep/wake stages classification using data driven methods", Biomedical Signal Processing and Control, vol. 2, 2007, pp. 171–179.
11. I. Zhovna, I. Shallom, "Multichannel Analysis of EEG Signal Applied to Sleep Stage Classification", Recent Advances in Biomedical Engineering, 2009, ISBN: 978-953-307-004-9.
12. N. Kerkeni, F. Alexandre, M.H. Bedoui, L. Bougrain, M. Dogui, "Automatic classifcation of sleep stages on a EEG signal by artificial neural networks", In proc. of the 5th WSEAS International Conference on Signal, Speech and Image Processing, WSEAS SSIP'05, 2005.
13. K.A.I. Aboalayon, H.T. Ocbagabir, M. Faezipour, "Efficient sleep stage classification based on EEG signals", In Proc. of the IEEE Systems, Applications and Technology Conference, LISAT, 2014.
14. U.R. Acharya, E.C. Chua, K.C. Chua, L.C. Min, T. Tamura, "Analysis and automatic identification of sleep stages using higher order spectra", Int. J. Neural Syst., vol. 20(6), 2010, pp. 509-21.
15. P.A. Estevez, C.M. Held, C.A. Holzmann, C.A. Perez, J.P. Perez, J. Heiss, M. Garrido, P. Peirano, "Polysomnographic pattern recognition for automated classification of sleep-waking states in infants", Med Biol Eng Comput., vol. 40(1), 2002, pp. 105-13.
16. A. Noviyanto, A.M. Arymurthy, "Sleep stages classification based on temporal pattern recogniton in neural network approach", In Proc. of IEEE World Congress on Computational Intelligence (WCCI), 2012.
17. M. Langkvist, L. Karlsson, A. Loutfi, "Sleep stage classification using unsupervised feature learning", Advances in Artificial Neural Systems, Vol. 2012.
18. B. Kemp, A.H. Zwinderman, B. Tuk, H.A.C. Kamphuisen, J.J.L. Oberye, "Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG", IEEE Transactions on Biomedical Engineering, vol. 47(9), 2000, pp. 1185-1194.
19. A.L. Goldberger, L.A.N. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.-K. Peng, H.E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals", Circulation, vol. 101(23), 2000 (June 13).
20. I. Mporas, V. Tsirka, E.I. Zacharaki, M. Koutroumanidis, M. Richardson, V. Megalooikonomou, "Seizure detection using EEG and ECG signals for computer-based monitoring, analysis and management of epileptic patients", Expert Systems with Applications, vol. 42(6), 2015, pp. 3227–3233.
21. I.H. Witten, E. Frank, "Data mining: practical machine learning tools and techniques", 2nd ed, Morgan-Kaufman Series of Data Management Systems, San Francisco: Elsevier, 2005.
22. C. Burges, "A tutorial on Support Vector Machines for Pattern Recognition", Data Mining and Knowledge Discovery, vol. 2(2), 1998, pp. 121-167.
23. M.H. Silber, S. Ancoli-Israel, M.H. Bonnet, S. Chokroverty, M.M. Grigg-Damberger et al., "The visual scoring of sleep in adults", J. of Clinical Sleep Medicine, vol. 3(2),2007, pp. 121–31.
24. H.H. Jasper. 1958. The ten-twenty electrode system of the International Federation. Electroencephalography and Clinical Neurophysiology, 371-375.