Comparison of Artificial Immune Clustering with Fuzzy c-means Clustering in the Sleep Stage Classification Problem

M. Dursun
Ankara University, GAMA Vocational School,
Ankara, TURKEY
mdursun@ankara.edu.tr

S. Ozsen
Selcuk University, Dept. of Electrical-Electronics Eng.
Konya, TURKEY
seral@selcuk.edu.tr

S. Gunes
Selcuk University, Dept. of Electrical-Electronics Eng.
Konya, TURKEY
sgunes@selcuk.edu.tr

S. Yosunkaya
Konya University, Faculty of Medicine, Sleep Laboratory
Konya, TURKEY
syosunkaya@selcuk.edu.tr

***Abstract*— Automatic Sleep Staging is an active field of research in sleep staging community. Many methods have been applied to get rid of the cumbersome of manual staging process. Even though very effective results were taken in some of these methods with respect to the classification accuracy, they are restricted either with their limited classification data or with lower number of classified stages like wake, sleepy and deep sleep. The accuracies obtained with methods for the classification of whole sleep stages are very low to apply in real sleep staging. One reason for this is the class imbalance in training data. Approximately half of one-night sleep consists of Non-REM2 stage while Wake, Non-REM1 and Non-REM3 stages are comparatively short duration. So, the used systems can converge to the characteristics of Non-REM2 stage. Taking equal amounts of data from each stage in training can be a solution for this but in this time a question arises: which samples should be picked from the each stage. Clustering schemes can play their roles for this question. In this study, we realized this clustering process with two methods: Fuzzy C-means Clustering (FCM) and Artificial Immune Clustering (AIC). We used 55 features that extracted from the sleep EEG, EOG and EMG signals of 8 subjects. We took a total of 300 data from each stage using FCM and AIC and classified these data with Artificial Neural Networks. The performances of the used clustering methods were compared on different number of features for which PCA was applied. The results showed that AIC was over-performed to FCM by obtaining a classification accuracy of 80.62% while this accuracy was 72.16% with FCM method used.**

Keywords-component; EEG; Sleep Stage; C-Means Clustering; Artificial Neural Networks;

I. INTRODUCTION

Sleep is an indispensible part of our daily life. A healthy and productive day begins with a healthy sleep. Unfortunately, sleep disorders have been affecting many of us in different manners. Whether the cause of these disorders is, the consequences can be severe. One enters a series of stages during his sleep and the quality of sleep depends on the number and order of these stages. The names of these aforementioned stages are: Wake, Non-REM1, Non-REM2, Non-REM3 and REM stages. Detection of any sleep disorder requires a correct staging of sleep. This staging process is done by a sleep expert through using some signals of subject like Electroencephalogram (EEG), Electrooculogram (EOG), Electromyogram (EMG) that recorded during his sleep with Polysomnography (PSG) device. Generally PSG recordings are divided into 10, 20, 30 or 60 seconds epochs and sleep expert determines the stage of these epochs by evaluating related signals and specific signal patterns. He does this according to the generally accepted rules of Rechtschaffen and Kales (RKS) [1]. But manual staging is a very tiring and subjective process. Thus, from the 1980's a search for automatic sleep scoring systems has begun C.A. Holzmann, C.A. Pérez, C.M. Held, M.San Martín, F. Pizarro, J.P. Pérez, M. Garrido and P. Pierano developed an expert system by using ganglionar lattices for infants and they obtained an overall performance of 96,4% agreement with the expert on validation data without artefacts and 84,9% agreement on validation data with artefacts [2]. E. Oropesa, H.L. Cycon and M. Jobert used Discrete Wavelet Transform to divide EEG signal into 7 specific frequency bands for sleep staging [3]. They reached an agreement level of 76,6%. In another study, Agarval and Gotman used the segmentation and clustering strategies in their application which also involves active participation of sleep operator [4]. They reached 80,6% classification accuracy in their study. E. Estrada, H. Nazeran, P. Nava, K. Behmehani, J. Burk and E. Lucas tried three different feature extraction schemes which were Relative Spectral Band Energy, Harmonic Parameters and Itıkura Distance [5]. Hidden Markov Models are one of the methods applied to this classification area [6]. A. Flexer, G. Gruber and G. Dorffner used two datasets from two labs and reached a classification accuracy of 54% and 42,02% for that labs respectively. Susmakova and Krakovska on the other hand examined 73 characteristics measures in sleep staging process with discriminant analysis by Fisher's quadratic classifier [7].

There are huge numbers of studies conducted in this regard but generally the performances of the used methods are low- in the orders of 70-85% intervals when five stages

978-1-4673-1448-0/12/$31.00 ©2012 IEEE

are used in classification. There are some reasons related with used methods and one of them is unbalanced data. A person in an 8-hour sleep enters Non-REM2 sleep approximately half of his sleep. Very few epochs are seen in Wake, Non-REM1 and NonREM3 sleep. Thus, the used classifier converges to the data in Non-REM2 sleep. To prevent this to occur, equal amounts of data from each stage can be taken for the training process. But in this time, another problem comes into existence: which data should be used in each stage. This is a general problem of clustering schemes and can be solved by them. Therefore, we used two clustering scheme in this study to have a balanced training set and compared their results on Artificial Neural Networks (ANN) classifier. The used clustering methods are: Fuzzy c-means clustering (FCM) and Artificial Immune Clustering (AIC) methods. EEG, EOG and EMG signals of 8 voluntary subjects were recorded in Meram Medicine Faculty of Konya University and 56 features were extracted from these signals. Approximately 300 data from each stage were taken by FCM and AIC respectively and an optimum ANN classifier was trained & tested with these data. In addition to this a feature reduction process was conducted with the aid of Principal Component Analysis (PCA) before the clustering schemes and by doing this performance comparisons of two clustering methods were conducted for different numbers of features. According to the obtained results at the end of experiments, AIC was reached a classification accuracy of 80.62% with 20 features used. The FCM methods on the other hand provided ANN to give a classification accuracy of 72.16% with 40 features. Besides of these, a training with unbalanced data was conducted to see the effect of used methods and a classification accuracy of 76.34% was recorded with optimum ANN used. However the result is seem to be moderate, the actual deficits were seen when the contingency tables were analyzed. The methods, related results and comments are given in the proceeding sections.

II. MATERIAL AND METHODS

A. Used dataset

The PSG recordings of 6 subjects which were recorded at the Meram Faculty of Medicine in Konya University were used during the experimentations. EEG, EMG and EOG signals of these subjects were first divided to 30-s epochs and a Doctor classified these epochs manually. The distribution of epochs according to the stages is given in Table I.

TABLE I. DISTRIBUTION OF EPOCHS WITH RESPECT TO THE SLEEP STAGES

Wake	Non-REM1	Non-REM2	Non-REM3	REM	TOTAL
627	398	3554	255	1012	5846

By using the EEG, EMG, left eye EOG and right eye EOG signals in each epoch, 52 features were extracted to use in the experiments which are:

Features 1-5: Mean, standard deviation, skewness, kurthosis and shape factor of time domain EEG signal.
Features 6-11: Absolute total powers of alpha, theta and delta band frequencies and 12-14 Hz, 2-6 Hz and 0.5-2 Hz frequency ranges in the frequency spectrum of EEG signal
Features 12-17: Relative powers of alpha, theta and delta band frequencies and 12-14 Hz, 2-6 Hz and 0.5-2 Hz frequency ranges in the frequency spectrum of EEG signal
Feature 18: Power of alpha band in related epoch/power of alpha band in previous epoch
Feature 19: Power of alpha band in related epoch/power of alpha band in next epoch
Feature 20: Power of theta band/power of alpha band
Features 21-22: Mean and maximum value of time domain EMG signal
Features 23-26: Sum, Maximum, Mean and standard deviation of Powers in the EMG frequency spectrum
Features 27-28: The absolute and relative energy of the time domain EMG signal in that epoch
Feature 29: The energy of EMG signal in time domain for that epoch/ for previous epoch
Feature 30: The energy of EMG signal in time domain for that epoch/ for next epoch
Features 31-37: Mean value, energy, maximum value, standard deviation, shape factor, skewness and kurthosis of time domain left eye EOG signal.
Feature 38-39: Absolute and relative powers of frequencies in 0.5-2 Hz in frequency spectrum of left eye EOG signal.
Feature 40: Mean value of the frequency spectrum of left eye EOG for that epoch /for all epochs
Features 41-47: same features with features 25-31 but in this time by using right eye EOG signal.
Feature 48: Mean value of the signal produced by summing left and right EOG signal
Feature 49: Energy of the signal produced by summing left and right EOG signal
Feature 50: Standard deviation of the signal produced by summing left and right EOG signal
Feature 51: Energy of summation signal in that epoch/ energy of the summation signal in previous epoch
Feature 52: Energy of the summation signal in that epoch/ energy of the summation signal in next epoch

B. Applied Method

After extracting features, FCM and AIC clustering methods were applied separately to take equal amounts of data from each stage. But before this process, principal component analysis (PCA) method was used to reduce number of features. The general procedure of the whole experimentation is shown in Figure 1. Here in the PCA feature reduction stage, number of features was reduced to 50, 40, 30, 20, 10, 8, 6, 4 and 2. Data reduction and ANN classification processes were realized for each of these feature numbers in both procedures. In data reduction stage, which was done with FCM and AIC separately, 300 data from Wake, Non-REM1, Non-REM2 and REM stages were taken. The number of data in Non-REM3 stage is 255. Thus no data reduction was conducted for that stage.

1) Data Reduction with FCM Method: Clustering algorithms have being used for such purposes of data reduction. Taking the centre of the cluster other than each member in that cluster can solve the problem. In FCM method, a fuzzy membership is assigned to each data point corresponding to each cluster centre according to their distances. After each iteration, membership and cluster centres are updated according to the formula [8]:

$$\mu_{ij} = 1 / \sum_{k=1}^{c} \left(d_{ij} / d_{ik} \right)^{(2/m-1)} \tag{1}$$

$$v_j = \left(\sum_{i=1}^{n} (\mu_{ij})^m x_i \right) / \left(\sum_{i=1}^{n} (\mu_{ij})^m \right), \forall j = 1, 2, ...c \tag{2}$$

where, *'n'* is the number of data points. *'vj'* represents the j^{th} cluster centre. *'m'* is the fuzziness index m € [1, ∞]. *'c'* represents the number of cluster centre. *'µij'* represents the membership of i^{th} data to j^{th} cluster centre. *'dij'* represents the Euclidean distance between i^{th} data and j^{th} cluster centre. Main objective of fuzzy c-means algorithm is to minimize:

$$J(U,V) = \sum_{i=1}^{n} \sum_{j=1}^{c} (\mu_{ij})^{m} \left\| x_i - v_j \right\|^2 \tag{3}$$

where, '$||x_i - v_j||$' is the Euclidean distance between i^{th} data and j^{th} cluster center.

2) Data Reduction with AIC Method: Like other areas, artificial immune systems have also being used as clustering methods effectively. The clonal selection principle is generally modelled in these types of methods and it can be summarized here as shortly [9]:

input : S = set of patterns to be recognised, n the number of worst elements to select for removal
output: M = set of memory detectors capable of classifying unseen patterns
begin
Create an initial random set of antibodies, A
forall patterns in S do
Determine the affinity with each antibody in A
Generate clones of a subset of the antibodies in A with the highest affinity. The number of clones for an antibody is proportional to its affinity
Mutate attributes of these clones inversely proportional to its affinity. Add these clones to the set A, and place a copy of the highest affinity antibodies in A into the memory set, M
Replace the n lowest affinity antibodies in A with new randomly generated antibodies
end
end

The generated memory antibodies in the above AIC method can be used as data to be taken from the larger input space. During the training with ANN, 70% of data in each stage was reserved for training while the remaining was used as test data. In each ANN training an ANN architecture of fnxhnx5 was utilized where fn is the feature number, hn is the hidden node number. Because there are 5 sleep stage, 5

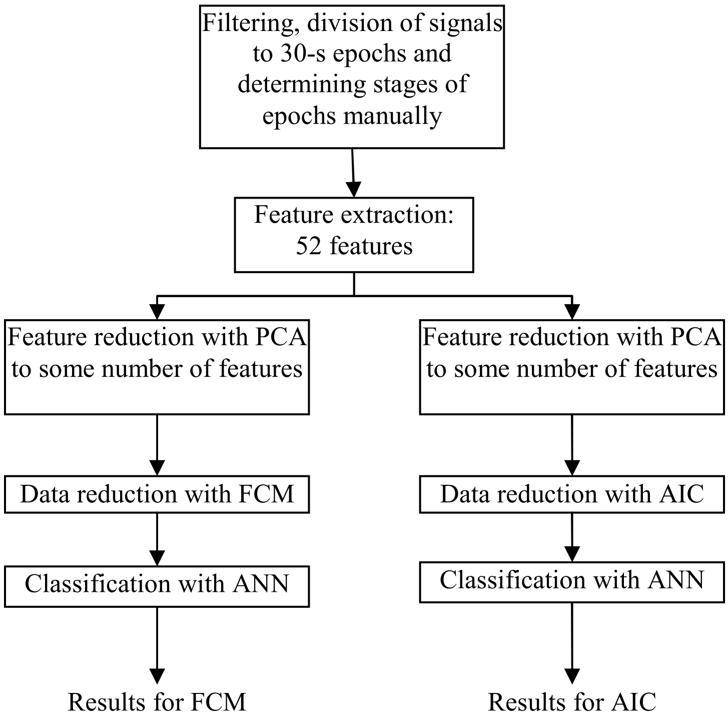

Figure 1. The block diagram of the experimental procedure

output nodes were used. While training an ANN, some optimization procedures should be conducted. Thus, we tried to find optimum ANN by doing some experiments. For example, we found optimum hn number by changing hn number from 1 to 80 and training ANN for each of them to find minimum test classification error. Other ANN parameters were found by similar manner.

III. RESULTS

As stated in the above section, feature selection with pca was conducted to obtain some number of features which were: 52 (no feature selection), 50, 40, 30, 20, 10, 8, 6, 4 and 2. After than 300 epochs from each stage (other than NREM3 which has 255 epochs) were extracted by using FCM and AIC methods respectively. For each experiment conducted for a definite feature number, optimum ANN configuration was searched (optimum number of node in hidden layer was searched). In Table II, the obtained results for each experimented feature number by using FCM and AIC data reduction methods are given.

TABLE II. THE RESULTS OF EXPERIMENTS CONDUCTED FOR EACH NUMBER OF FEATURES BY USING FCM AND AIC METHOD

Feature number	**Optimum *hn* for FCM**	**Min. Classification error of FCM (%)**	**Optimum *hn* for AIC**	**Min. Classification error of AIC**
52 *	38	28.04	41	25.12
50	32	27.83	38	23.06
40	80	27.83	45	24.12
30	52	28.04	16	23.05
20	32	32.37	21	22.16
10	18	38.55	12	26.85
8	28	40.00	63	29.15
6	24	41.44	41	30.37
4	52	40.41	70	32.46
2	70	50.92	65	40.41

a. * no feature selection

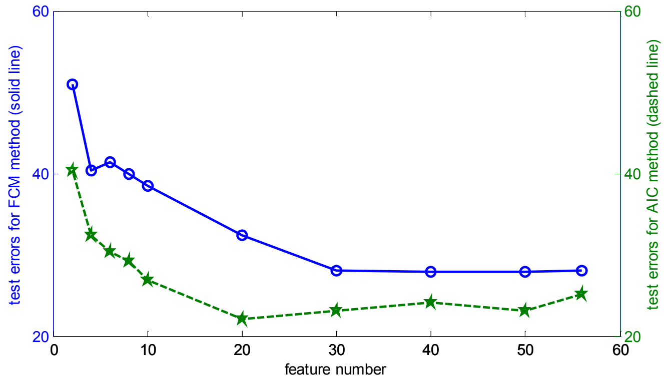

Figure 2. Change of test classification error with respect to the selected feature number for FCM and AIC methods.

The change of test classification error with respect to the feature number is shown for both FCM and AIS methods in Figure 2.

As shown from the Figure 2 and Table II, AIC was over-performed to FCM method for every feature numbers. The minimum test classification error with FCM method was obtained as 27.83% for feature numbers 40 and 50. This result was recorded as 22.16% for AIC method. But these results were only found by changing hn number in ANN configuration. For that reason, after determining optimum feature number and hn number for FCM and AIC methods, optimum value of iteration number, learning parameter and momentum constant are searched. The following results (look at TABLE III.) were taken in this search for two methods:

TABLE III. THE RESULTS OF EXPERIMENTS FOR TWO METHODS

	FCM	**AIC**
lr	2.6	2.8
mc	0.8	0.85
It_num	5670	4170
Feature number	40	20
hn	80	16

According to the results, AIC method has over-performed to FMC method by an increase in correct classification ratio of 6.74%.

IV. DISCUSSION

In this study, we compared the data reduction capabilities of FCM and AIC methods on the sleep stage classification problem. For this purpose we extracted 52 features from the time and frequency domain EEG, EMG and EOG signals. After obtaining different numbers of features from this feature set by PCA, we get 300 features from each stage by using FCM and AIC methods separately. To compare these reduction processes we tried to find best ANN configuration through experimenting ANN for optimum hidden node number, learning rate, momentum constant and maximum iteration number.

According to the results, it was found that a classification accuracy of 73.88% was obtained for 40 features when FCM data reduction method was used. On the other hand, AIC method has reached a classification accuracy of 80.62% for only 20 features. That is, it can be said that, in addition to the advantage in obtaining a higher classification accuracy, an advantage with respect to the time consuming can also be regarded to the AIC method.

REFERENCES

[1] A. Rechtschaffen and A. Kales, A Manual of Standardized Terminology. Techniques and Scoring System for Sleep Stages of Human Subjects, US Government Printing Office, Washington, 1969.

[2] C.A. Holzmann, C.A. Pérez, C.M. Held, M.San Martín, F. Pizarro, J.P. Pérez, M. Garrido, and P. Pierano, “Expert-system classification of sleep/waking states in infants,” Medical & Biological Engineering and Computing, vol. 37, no. 4, pp. 466-476, 1999.

[3] E. Oropesa, H.L. Cycon, and M. Jobert, Sleep stage classification using wavelet transform and neural network, International Computer Science Institute, TR-99-008, March1999.

[4] R. Agarwal and J. Gotman, “Computer-Assisted Sleep Staging,” IEEE Transactions on Biomedical Engineering, vol. 48, no. 12, pp. 1412-1423, 2001.

[5] E. Estrada, H. Nazeran, P. Nava, K. Behmehani, J. Burk, and E. Lucas, “EEG feature extraction for classification of sleep stages,” 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2004, IEMBS '04, vol.1, pp. 196-199, 2004.

[6] A. Flexer, G. Gruber, and G. Dorffner, “A reliable probabilistic sleep stager based on a single EEG signal,” Artificial Intelligence in Medicine, vol. 33, no. 3, pp. 199-207, 2005.

[7] K. Šušmáková and A. Krakovská, “Discrimination ability of individual measures used in sleep stages classification,” Artificial Intelligence in Medicine, vol. 44, no. 3, pp. 261-277, 2008.

[8] S. Theodoridis & K. Koutroumbas, Pattern Recognition, USA, 2003.

[9] L. N. de Castro and J. Timmis, Artificial Immune Systems: A New Computational Intelligence Approach, 1st ed., Heidelberg, Germany, Springer-Verlag,2002.