# A Comparative Study between SVM and Fuzzy Inference System for the Automatic Prediction of Sleep Stages and the Assessment of Sleep Quality

Ch. Panagiotou(*), I. Samaras, J. Gialelis, P. Chondros, D. Karadimas
Electr. & Comput. Eng. Dept. University of Patras, Greece
(*)Corresponding author email: cpanag@ece.upatras.gr

**Abstract—this paper compares two supervised learning algorithms for predicting the sleep stages based on the human brain activity. The first step of the presented work regards feature extraction from real human electroencephalography (EEG) data together with its corresponding sleep stages that are utilized for training a support vector machine (SVM), and a fuzzy inference system (FIS) algorithm. Then, the trained algorithms are used to predict the sleep stages of real human patients. Extended comparison results are demonstrated which indicate that both classifiers could be utilized as a basis for an unobtrusive sleep quality assessment.**

***Keywords—EEG, sleep stages, SVM, FIS***

## I. INTRODUCTION

The relevance of sleep abnormalities with chronic diseases and inflammatory conditions such as depression, heart disease, obesity, diabetes, stroke and arthritis has been already manifested in the literature [1-3]. Currently, there is a significant number of people suffering from sleep disorders, like insomnia, narcolepsy, sleep apnea, etc. The constantly increasing number of people facing sleep abnormalities intensifies the strong relation between sleep quality and quality of life [4-5].

Polysomnography test has been during the past years the dominant tool for sleep quality monitoring and assessment. However, this test requires specialized equipment and can be performed only within a certified lab. Thus, patients cannot repeat this test routinely and in their convenience (i.e. at their home environment). The appropriate, non–obtrusive way for assessing the sleep quality refers to subjective metrics and methods such as the Pittsburgh questionnaire (PSQI)[6].

Sleep is a dynamic phenomenon which is characterized by individual sleep stages. These sleep stages alter during person's rest sessions and contribute towards his sleep cycle formation. The two main sleep stages are the Random Eye Movement stage (REM) and the non – REM stage (NREM). More precisely, NREM can be further distinguished in 4 stages. Stages 1 and 2 are denoted as light sleep phase while stages 3 and 4 as deep sleep or slow waves phase. These stages and the related brain wave frequencies are presented in Table I. The transitions from stage to stage and the duration of each stage during a person's sleep are the main markers for sleep quality assessment.

According to literature quantitative analysis of sleep electroencephalography (EEG) data can provide valuable additional information in sleep research [7].These data recordings are considered a reliable method of assessing a person's sleep stages. However, recent evolution in artificial intelligence has encouraged new efforts on the detection of sleep stages and finally the assessment of sleep quality through machine learning algorithms. Therefore, clinicians and researchers' effort has nowadays been focused on analyzing and extracting enriched features that could feed classifiers, in order to produce efficient and accurate models for the identification of each person's sleep cycle.

TABLE I. SLEEP STAGES AND BRAIN WAVES

| Stage | Frequency (Hz) | Brain Wave |
|---|---|---|
| Awake | 13-30 | Beta |
| 1 | 8-13 | Alpha |
| 2 | 4-13 | Spindle and theta waves |
| 3 | 2-4 | Spindle and Delta waves |
| 4 | 0.3-2 | Delta waves |
| REM | 13-30 | |

This paper is organized as follows. Section II demonstrates the related work and Section III analyzes the methodology employed in this study. The experimental results obtained by applying two different machine learning algorithms are presented in Section V while Section VI concludes the paper.

## II. RELATED WORK

A performance comparison among popular classifiers for the detection of sleep stages is presented is presented in [8]. More precisely, SVM ensemble and Random forest has been tested on ten healthy subjects. In this study the random forest algorithm fed with spectral linear features has outperformed SVM.

An artificial neural network approach reaching 76% performance of identifying stages 1,2,3,4, REM and wake is presented in [9]. The reformation of stages in three larger groups such as (wake), (stage 1, stage 2, REM), (stage 3 stage 4) increased the performance by 82%.

The authors of [10] present another comparison of sleep stage classification by testing the performance of k – Nearest

PervasiveHealth 2015, May 20-23, Istanbul, Turkey

DOI 10.4108/icst.pervasivehealth.2015.259248

Neighbor (kNN), Quadratic Discrimminant Analysis and SVM. In these experiments the SVM achieved the most accurate classification by identifying correctly the 73.1% of the stages on healthy subjects and 76% on subject with obstructive apnea.

Finally in [11] an SVM classifier is applied on the proposed features that derive from detrended fluctuation analysis on the ECG (MIT – BIH polysomnography database), achieving a classification rate of 80%.

However the clinical golden standard so far has been the manual scoring from medical experts while the applications and devices that estimate sleep quality indices, based on actigraphy, have not been proven yet as reliable enough to produce accurate and significant outcomes.

In the current work we present the main guidelines for a sleep stage prediction system used as a primary screening and unobtrusive tool for sleep quality assessment. For this purpose, we have compared two dominant machine learning algorithms, namely the support vector machines (SVM) and the fuzzy inference systems (FIS). Both of them present significant advantages and tradeoffs. While many prediction techniques have been reviewed side – by – side for the sleep stage classification problem, a straightforward comparison among FIS and SVM on the same basis has not been observed. A feasibility study concerning the reliable sleep stage assessment using these two algorithms is the main aim of the current study.

## III. METHODOLOGY

### A. *Data*

The extraction of reliable and accurate models based on data mining and machine learning techniques requires numerous datasets. However, data mining and machine learning in the healthcare domain lacks of data availability. On that context public databases, that make available medical data, try to address this issue. For our study the MIT-BIH Posylomnographic Database [12] which is available from Physionet [13] has been employed. This database contains physiological signals that were recorded during the sleep session of 16 subjects. In total the database consists of 18 records with over 80 hours of polysomnographic recordings. The recorded signals are the electrocardiogram (ECG), EEG, Electroocculogram (EOG) and respiration rate. Since our effort focuses on the development of a non - obtrusive system, the usage of a single physiological signal was a demanding specification. Thus, the EEG signal was considered as the main signal that manifests the sleep activity. EEG records have been recorded with a sampling rate of 250Hz and have been annotated by medical experts every 7500 samples (30 seconds).

### B. *Feature extraction*

The brain activity is captured in the EEG signals as voltage alterations that hardly exceed a threshold of 100μV. The low amplitudes of an EEG signal are prone to increased signal – to – noise ratio. EEG is highly affected by surrounding signals such as body movement, eye blinking, ECG and the power line inference. Therefore a preprocessing step is required for the extraction of these artifacts from the EEG. On the employed EEG signals a band pass filter with cutoff frequencies at 0.3Hz and 40Hz has been applied.

In order to extract meaningful and semantic information from the EEG signal a further processing stage was conducted. This stage produced features that have been extracted from the frequency and the time domain. Since the sleep stages are strongly related with the brain waves (presented in Table I) the respective frequency bands and their power spectrum have been extracted with a 512 samples Hanning window and 50% overlap. Some further statistic calculations have been applied for the extraction of the frequency, with the higher power on each epoch, and the median frequency spectral power.

Statistical and time domain analysis has been proven to extract useful characteristics that expose important patterns of the brain activity. Based on a thorough study of previous works [14], [15] we extracted the time domain features that has been proved to present the higher degree of correlation with sleep activity. This processing resulted in the extraction of the respective following features:

- Hjorth Mobility & Complecity
- Kurtosis & Skewness
- Interquartile range
- Maximum, minimum, mean and range
- Variance standard deviation
- Shannon Entropy
- Zero Crossing Points (strongly related with presence of spindles)
- Mean absolute and median absolute deviation

Finally our datasets have been completed with the auto regressive filter coefficients extracted from the 6$^{th}$ order auto regression analysis of the signals. All the processing and feature extraction has been applied on non – overlapping EEG epochs of 30 seconds duration.

### C. *Classification Strategy*

For the final sleep stage prediction, we have employed classification techniques from the machine learning domain, as it has been already mentioned. Especially we focused on a performance comparison between two well established and popular techniques of supervised learning, namely the SVM and the FIS.

The cornerstone of every machine learning classification approach are the data that feed and train the classifiers. However, medical data present very low availability for the researchers. The datasets constructed from the MIT – BIH sleep database offer a sufficient quantity but the classes are not uniformly distributed so as to construct an ideal and unbiased dataset. In total, the dataset consisted of 10181 instances (each instance refer to on 30 seconds EEG epoch) with the following distribution of classes:

TABLE II. CLASSES DISTRIBUTION

| Groups | Sleep Stages | # Stages | # Groups |
|---|---|---|---|
| Wake | Wake | 3120 | 3120 |
| Light Sleep | REM | 700 | 6397 |
| | Stage 1 | 1814 | |
| | Stage 2 | 3883 | |
| Deep Sleep | Stage 3 | 483 | 664 |
| | Stage 4 | 181 | |
| Total: | | 10181 | 10181 |

In order to overcome such problems that cause overfitting issues we used the k-fold Cross Validation technique for evaluating the classifiers. In k-fold cross validation the training set is randomly divided into K disjoint sets of equal size with similar class distribution in every set. Then the classifier is trained with the respective k-th training set while its performance is evaluated with the respective test set that was held out. Finally the estimated performance metric is the average of the values obtained from the k folds. A second approach we followed in order to increase as possible the number of instances from each class was to group them in three classes by joining the classes with common characteristics (clinical, qualitative and quantitative). For example Stage 3 and 4 both share the presence of dominant slow waves, while in stages 1 and 2 are occur theta waves.

The classifiers' performance have been evaluated from the study of the respective confusion matrices derived from the experiments along with the accuracy metric, defined as: $accuracy = \frac{correctly\_identified\_records}{total\_records}$ . Further popular metrics used for the evaluation of the classification performance of multiclass classifiers are recall (rec) and precision (prec) defined as: $rec = \frac{TP}{TP+FN}$, $prec = \frac{TP}{FP+TP}$, where TP: True Positive, TN: True Negative and FP: False Positive (respective metrics for binary classification are sensitivity and specificity)[16].

*1) FIS*

A fuzzy rule-based expert system contains fuzzy rules in its knowledge base and derives conclusions as outputs from the user inputs and the fuzzy reasoning process. All these features constitute a fuzzy inference system (FIS) [17].

In this study, the learning algorithm introduced in [17] was used in order to automatically derive the membership functions and the fuzzy IF/THEN rules from the real EEG data together with its corresponding sleep stages.

In particular, initially the subtractive clustering algorithm [17] was utilized for separating the training EEG data together with its corresponding sleep stages into clusters. This algorithm does not involve any iterative nonlinear optimization, and therefore is robust and fast. The following value was defined for each training instance $P_s = \sum_{j=1}^{50} e^{-a\|L_s - L_j\|^2}$ where $s \in [1,50]$, $a$ is a positive constant (here $a = 0.5$), and $L_s$ denotes the multidimensional real numerical data of the $s$th training instance. Then, the procedure described in [5] has been utilized. The constructed FIS parameters are listed in Table II.

*2) SVM*

SVM algorithms have offered great impact on the evolvement and the application of machine learning in general. SVM consider the data points as vectors in a high dimensional space and tries to estimate the optimal hyperplane that separates the data in the respective classes. This properties describe SVM as a binary linear classifier. The linearity however can be overridden through the adoption of the kernel methods instead of vectors. Kernel methods maps the data points of the training sets to hyperplanes that may offer better separation among the classes of the data.

TABLE II. PARAMETERS OF THE FIS.

| **Parameter** | **Value** |
|---|---|
| AND method | Algebraic product |
| OR method | Probabilistic OR |
| Implication method | Algebraic product |
| Aggregation method | Max |
| Type of membership functions | Gaussian |
| Fuzzy inference method | Sugeno |
| Defuzzifier | Weighted average |

SVM present high degree of generalization performance in many problems but they add also significant computational complexity during the training phase. This drawback has been partially addressed in [18] with the Sequential Minimal Optimization that we employ in our study along with the polynomial kernel $K(x,y) = (x^T + c)^d$ where x and y are vectors of the feature space and c a constant and d a positive integer.

## IV. EXPERIMENTAL RESULTS

As already described our evaluation strategy was based on the k – fold Cross Validation (k = 10) on a multiclass dataset with 3 classes (Wake), (S1, S2, REM) and (S4, S5). The classifiers were tested initially on the whole dataset with the 10181 instances. The confusion matrix for the FIS and SVM classifiers are depicted on Table III and Table IV respectively.

TABLE III. FIS CONFUSION MATRIX WITH UNBALANCED DATASET

| **Classified as:** | **W** | **(S1,S2,REM)** | **(S3,S4)** | **Total** |
|---|---|---|---|---|
| W | **2160** | 940 | 20 | 3120 |
| (S1,S2,REM) | 1076 | **5304** | 7 | 6397 |
| (S3,S4) | 30 | 264 | **370** | 664 |
| **Total** | 2980 | 6774 | 427 | 10181 |
| **Performance Metrics** | rec=69% prec=72% | rec=83% prec=78% | rec=56% prec=87% | Acc= **77%** |

TABLE IV. SVM CONFUSION MATRIX WITH UNBALANCED DATASET

| **Classified as:** | **W** | **(S1,S2,REM)** | **(S3,S4)** | **Total** |
|---|---|---|---|---|
| W | **2147** | 966 | 7 | 3120 |
| (S1,S2,REM) | 487 | **5775** | 135 | 6397 |
| (S3,S4) | 7 | 233 | **424** | 664 |
| **Total** | 2641 | 6974 | 566 | 10181 |
| **Performance Metrics** | rec=69% prec=81% | rec=90% prec=83% | rec=64% prec=75% | Acc=**82%** |

From this first round of experiments we observe that we get a classification accuracy of 82% for the SVM classifier while the FIS reaches the 74%. From a more thorough study of the two confusion matrices we observe how the unbalanced training dataset is highlighted on the classification stats for each class. The FIS identifies the (S1, S2, REM) class with 83% success rate while the SVM with 90% (respective recall metrics), exceeding both their average hit rate (accuracy). Significant poor performance is also observed in the identification of deep sleep stages. This is obviously attributed to the biased dataset, since the instances labeled as (S1, S2, and REM) and (S3, S4) are the 62% and 6.5% respectively of all instances.

The next experiments were conducted with balanced datasets in order to study how the distribution of classes in the training data affect the performance of the classifiers. The class containing the less instances is the deep sleep stage (S3, S4) which consists of 664 instances. In order to construct the balanced dataset we selected randomly 664 instances from the two other sets of instances labeled as (W) and (S1, S2, REM) respectively. The new dataset now consists of a total number of 3*664 = 1992 instances. The tradeoff for building the balanced dataset is that now we utilize only the 20% of the available instances. The experiments have been repeated multiple times after selecting randomly 664 instances from the classes with a surplus of instances. The respective statistic results from the classification processes are presented on Table V and Table VI for FIS and SVM respectively.

TABLE V. FIS CONFUSION MATRIX WITH BALANCED DATASET

| **Classified as:** | **W** | **(S1,S2,REM)** | **(S3,S4)** | **Total** |
|---|---|---|---|---|
| W | **500** | 159 | 5 | 664 |
| (S1,S2,REM) | 145 | **450** | 69 | 664 |
| (S3,S4) | 11 | 80 | **573** | 664 |
| **Total** | 632 | 682 | 678 | 1992 |
| **Performance Metrics** | rec=75% prec=79% | rec=68% prec=66% | rec=86% prec=85% | **Acc**=78% |

TABLE VI. SVM CONFUSION MATRIX WITH UNBALANCED DATASET

| **Classified as:** | **W** | **(S1,S2,REM)** | **(S3,S4)** | **Total** |
|---|---|---|---|---|
| W | **541** | 118 | 5 | 664 |
| (S1,S2,REM) | 80 | **515** | 69 | 664 |
| (S3,S4) | 11 | 49 | **604** | 664 |
| **Total** | 632 | 682 | 678 | 1992 |
| **Performance Metrics** | rec=81% prec=86% | rec=78% prec=76% | rec=91% prec=89% | **Acc**=83% |

None of the two classifiers present significant improvement on the measured accuracy after the completion of the second configuration of the experiments. However we observe as expected a more balanced performance on the identification on each particular class. Particularly now the deep sleep stage is identified with accuracy 86% and 91% for the FIS and SVM respectively. This can be attributed to the fact that sleep stages 3 and 4 that form the deep sleep stage have characteristics that do not overlap with any of the other classes (i.e. the Delta waves). In contradiction the early sleep stage 1 which is assigned to the light sleep stage is harder to de identified from REM and wake.

Both classifiers achieved satisfactory results but they present potentials for further improvement. This improvement could be achieved from the enhancement of the dataset with new features either from the dimensionality reduction of the dataset through sophisticated feature selection algorithms. The SVM achieved significant accuracy over 80% but the FIS gave us a 78% with less computational complexity.

## V. CONCLUSION

In this paper, two well-known machine learning algorithms, namely SVM and the FIS, have been used for the prediction of the human sleep stages. These algorithms have been trained with real human EEG data together with corresponding sleep stages. The trained algorithms have been assessed in cases of predicting the sleep stages of real human patients. This assessment has shown that the SVM verified the expectations for better performance over the FIS, but both techniques can deliver sufficient accuracy. The quality of life for people suffering from chronic diseases and sleep disorders could be benefitted by tools that monitor and assess their sleep.

## ACKNOWLEDGMENT

This work was financially supported by the Operational Program DEPIN 2007-2013 of Western Greece Region under the LEADERA framework, project “Healthy Sleep and Exercise Analysis Tool (hSEAT).

## REFERENCES

[1] Von Ruesten, Anne, et al. "Association of sleep duration with chronic diseases in the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam study." PLoS One 7.1 (2012): e30972.

[2] Foley, Daniel, et al. "Sleep disturbances and chronic disease in older adults: results of the 2003 National Sleep Foundation Sleep in America Survey." Journal of psychosomatic research 56.5 (2004): 497-502.

[3] Ranjbaran, Z., et al. "The relevance of sleep abnormalities to chronic inflammatory conditions." Inflammation Research 56.2 (2007): 51-57.

[4] Albuquerque, Rachel et al.. "Sleep quality implicates in life quality: An analysis about children who suffered burns." Burns 40.4 (2014): 774-775.

[5] Do Valle, C. Ribeiro, et al. "Quality of life and sleep disorders in elderly." Sleep Medicine 14 (2013): e291.

[6] Shahid, Azmeh, et al. "Pittsburgh Sleep Quality Index (PSQI)." STOP, THAT and One Hundred Other Sleep Scales.Springer New York,2012. 279-283.

[7] P. Anderer, et al., "Artifact processing in computerized analysis of sleep EEG - a review", *Neuropsychobiology*, vol.40,no.3,Sep 1999,pp.150-157.

[8] Radha, Mustafa, et al. "Comparison of feature and classifier algorithms for online automatic sleep staging based on a single EEG signal." Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE. IEEE, 2014.

[9] Ronzhina, Marina, et al. "Sleep scoring using artificial neural networks." *Sleep Medicine Reviews* 16.3 (2012): 251-263.

[10] Yılmaz, Bülent, et al. "Sleep stage and obstructive apneaic epoch classification using single-lead ECG." *Biomedical engineering online* 9 (2010): 39.

[11] Adnane, Mouradet al. "Sleep–wake stages classification and sleep efficiency estimation using single-lead electrocardiogram." *Expert Systems with Applications* 39.1 (2012): 1401-1413.

[12] Lee, Jong-Min, et al. "Detrended fluctuation analysis of EEG in sleep apnea using MIT/BIH polysomnography data." Computers in Biology and Medicine 32.1 (2002): 37-47.

[13] Goldberger, Ary L., et al. "Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals." Circulation 101.23 (2000): e215-e220.

[14] Mporas, Iosif, et al. "Online seizure detection from eeg and ecg signals for monitoring of epileptic patients." Artificial Intelligence: Methods and Applications. Springer International Publishing, 2014. 442-447.

[15] Koley, B., and D. Dey. "An ensemble system for automatic sleep stage classification using single channel EEG signal." Computers in biology and medicine 42.12 (2012): 1186-1195.

[16] Sokolova, Marina, and Guy Lapalme. "A systematic analysis of performance measures for classification tasks." Information Processing & Management 45.4 (2009): 427-437.

[17] T.-P. Hong, and C.-Y. Lee, “Induction of fuzzy rules and membership functions from training examples,” Fuzzy Sets and Syst., vol. 84, no. 1, Nov. 1996, pp. 33-47..

[18] S. L. Chiu, “Fuzzy model identification based on cluster estimation,” J. Intelligent and Fuzzy Syst., vol. 2, no. 3, 1994, pp. 267-278.

Platt, John. "Fast training of support vector machines using sequential minimal optimization." Advances in kernel methods—support vector learning 3 (1999).