Ensemble SVM Method for Automatic Sleep Stage Classification

Emina Alickovic and Abdulhamit Subasi

Tweet

LINKÖPING UNIVERSITY

Ensemble SVM Method for Automatic Sleep Stage Classification

Emina Alickovic, Abdulhamit Subasi

Abstract— **Sleep scoring is used as a diagnostic technique in the diagnosis and treatment of sleep disorders. Automated sleep scoring is crucial, since the large volume of data should be analyzed visually by the sleep specialists which is burdensome, time-consuming tedious, subjective, and error-prone. Therefore, automated sleep stage classification is crucial step in sleep research and sleep disorder diagnosis. In the present article, a robust system, consisting of three modules, is proposed for automated classification of sleep stages from single channel EEG. In the first module, signals taken from Pz-Oz electrode were denoised using multiscale principal component analysis. In the second module, the most informative features are extracted using discrete wavelet transform (DWT) and then, statistical values of DWT sub-bands are calculated. In the third module, extracted features were fed into an ensemble classifier, which can be called as rotational support vector machine (RotSVM). The proposed classifier combines advantages of the principal component analysis and SVM to improve classification performances of the traditional SVM. The sensitivity and accuracy values across all subjects were 84.46% and 91.1% respectively for five stage sleep classification with Cohen's kappa coefficient of 0.88. Obtained classification performance results indicate that, it is possible to have an efficient sleep monitoring system with a single channel EEG, and can be used effectively in medical and home-care application.**

Index Terms—Sleep stage classification, single-channel EEG, multiscale principal component analysis (MSPCA), discrete wavelet transform (DWT), rotational support vector machine (RotSVM).

I. INTRODUCTION

THIS study presents a novel automated system for sleep stage classification which utilizes a single EEG channel. It is known that human beings spend around one third of their lives, in average, in sleep. Even though body physical activities are covert to a great extent, internal brain activities during sleep have power which cannot be easily understood and explained. A set of highly complex patterns happen in human brain during sleep. Recently, great amount of effort has been invested in sleep analysis studies and its connection to other psychological states. Yet, very little is known about sleep. Unfortunately, number of subjects suffering from sleep disorders is significant and it hardens subjects' everyday activities, besides it effects subject's health conditions in

Emina Alickovic is with the Department of Electrical Engineering, Linkoping University, Linkoping, 58183, Sweden. (e-mail: emina.alickovic@liu.se).

Abdulhamit Subasi is with the Computer Science Department, College of Engineering, Effat University, Jeddah, 21478, Saudi Arabia (e-mail: absubasi@effatuniversity.edu.sa).

different ways to great extent. Recent studies hypothesized that sleep may have significant task in memory consolidation where certain memories are build up while other less significant memories are vanished [1, 2]. Therefore, it is of great importance to have an accurate system for sleep monitoring and analysis of sleep behavior.

Objective sleep monitoring and analysis is commonly performed by expert(s) observing different sleep stages during whole night. As of the date, polysomnography (PSG) is "paragon of excellence" in sleep analysis. PSG can be understood as multivariate system which records different biological signals, like electroencephalogram (EEG), electrocardiogram (ECG), electromyogram (EMG), electrooculogram (EOG), concurrently. Once biological signals are collected, next step would be to agree on how collected signal should be scored. Rechtschaffen and Kales (R & K) [3] proposed a guide for classification of sleep which later became a golden standard and is employed as a tool for classification of sleep stages in numerous labs worldwide despite to its weaknesses [4]. R & K standard was further improved by American Academy of Sleep Medicine (AASM) [5]. Accordingly, sleep scoring distinguishes wakefulness (W) and one of two sleep stages: rapid eye movement (REM) and non-rapid eye movement (NREM). NREM can be further divided into 4 different stages enumerated as 1, 2, 3 and 4. Often NREM3 and NREM 4 are combined into one sleep stage called as slow-wave sleep (SWS). Yet, there could still be certain disagreement in manual scoring of sleep stages caused by experts' biased decisions and training education. Therefore, there is a need to develop an objective, non-biased automated sleep scoring system.

Many studies have been conducted with the aim to describe and detect different sleep stages [6, 7, 8, 9, 10]. In general, any objective, non-biased automated classification system consists of three different modules, namely pre-processing module, feature extraction module and classification module. In pre-processing module, normally noise and undesired signal components are removed using different filtering techniques. This can be achieved by detrending i.e. removing linear trends from signals and filtering out undesired frequency components. At this module, also different blind source separation algorithms, such as principal component analysis (PCA), independent component analysis (ICA), denoising source separation (DSS) etc., can be applied in order to clean the source signals from noise. Multiscale PCA (MSPCA) was proposed by Bakshi [11] to merge the capabilities of wavelet

transform (WT) and PCA. The PCA extracts the relationship between multiple variables, whereas the WT decorrelate the auto-correlation among the measurements [11]. MSPCA was already successfully applied for denoising different biomedical signals, e.g. EEG [12], EMG [13] and ECG [14] where significant improvement in classification accuracy was achieved. After removing undesired signal components, next step would be to extract the most informative features from signals. In literature, a wide variety of different approaches on how these features can be extracted are reported [12, 14, 13]. There is no rule which technique should be selected, but it depends on the nature and dynamics of signal.

In literature, different techniques are proposed for extraction of informative features from EEG signals that can reliably catch sleep dynamics. These include time-domain statistics approaches [15], spectral analysis [16], time-frequency analysis such as Wigner–Ville distribution [17] and wavelet analysis [7] and dynamic warping [18], graph domain analysis [9], coherence [19], etc. After it is decided on which features are to be extracted from the signals, next step would be to classify different states (such as wakeful vs sleep, or wakeful vs REM vs NREM, etc.). For this step, different machine learning (ML) techniques can be applied and there is a wide variety of different techniques reported in literature and how they can be applied for classification. They range from simple linear discriminant analysis (LDA) to non-linear and highly complex Gaussian mixture model-based classifiers [20, 21]. For classification of different sleep states, several traditional methods, namely LDA [18, 22], neural networks [23, 24], support vector machines (SVM) [25, 9, 26, 27, 6], *k*-nearest neighbor (k-NN) [16], hidden Markov model [28], fuzzy systems [29, 30], etc., are proposed for distinguishing between different sleep stages. The common to all traditional classifiers is that they have only one classifier. Recently, ensemble ML (EML) classifiers, where multiple traditional classifiers are combined, have been proposed to improve the performance of single classifier. One of the most applied EML methods, that found application in variety of research areas, is the random forest (RF) proposed by Breiman [31]. RF has also been proposed for identification of different sleep stages in [7].

The aforementioned techniques and proposed systems generally combine the features from different signal types (EEG, ECG, EMG and EOG) and perform classification with such features as inputs to the classifier. Perhaps more importantly, these systems can be further improved in terms of overall performances by proposing novel model for sleep stage classification that will use single-channel EEG signals while maintaining high classification performances. The contribution of this study is to use MSPCA for denoising and Rotational SVM for classification to create a reliable and efficient automated system for sleep stage identification and classification where the features will be extracted from a single-channel EEG signals. After segmenting Pz-Oz EEG channel signals, MSPCA is used to denoise the EEG signals in the pre-processing module. After denoising the EEG signals, in the second module, informative features from the denoised

signals are extracted using discrete wavelet transform (DWT), since it can efficiently decompose EEG signal into different frequency bands relevant to this study: delta (0.5-3 Hz), theta (4-7 Hz), alpha (8-12 Hz), beta (13-30 Hz) and low-gamma (30-50 Hz). Furthermore, in order to reduce the dimension of data, statistical values of DWT sub-bands are calculated to represent the distribution of wavelet coefficients in a better way. The extracted features are fed into the classifier in the third module. In this study, we propose a modified SVM which is called rotational SVM (RotSVM). The experimental results showed that it outperformed the results reported in similar studies done before.

The remainder of the paper is organized as follows. The next section describes the experimental data used in this study. In section III, methodology used for construction of single-channel sleep stage system together with theoretical background is explained. In section IV, experimental results obtained, when proposed system is applied on whole-night recorded sleep data, are reported. Afterwards, this section gives comparison of results obtained in this study with other results reported in literature. This section is concluded with discussion. Section V gives conclusion of this study.

II. EXPERIMENTAL DATA

A. Subjects and Data Collection

Data used in this study to evaluate the performances of proposed system was obtained from the Sleep EDF [Expanded] database [32, 33] which is publicly available online from Physionet Bank [34, 35]. This database is collection of whole-night PSG sleep records that include EEG (2 channels, Fpz-Cz and Pz-Oz), one horizontal EOG and EMG signal records together with their hypnograms (annotation of different sleeps stages). Sampling frequency for EEG signals was $F_s = 100\,\text{Hz}$. Since the focus of this study is to construct the system that will utilized only one EEG channel, Pz-Oz channel was selected since several recent studies reported that this channel provides higher classification performances [36, 22, 9]. Data contained in this database comes from two different studies where one study was conducted to understand age effects on sleep in healthy subjects (SC group) and contains two PSGs, each with duration around 20 hours whereas the second study was conducted to understand temazepam effects on sleep in Caucasian subjects (ST group) and contains one PSG, with duration of approximately 9 hours. In this study, 20 different subjects were considered where 10 subjects were randomly selected from SC groups and the remaining 10 subjects from ST group and in total we had 30 PSG records $(10\,subjects \cdot 2\,PSGs + 10\,subjects \cdot 1\,PSG = 30\,PSGs)$.

B. Manual Sleep Stage Scoring

Each PSG file in this database is associate with its corresponding hypnograms which were manually labeled according to R & K rules [3] by two well-trained sleep experts who labeled sleeps states independently, but according to Fpz-Cz/Pz-Oz EEGs instead of C4-A1/C3-A2 EEGs by proposed

by [37]. Whole night EEG signals were divided into 30 seconds long segments based on R & K suggestions [3]. Sleep stages were labeled as wakefulness (W), REM, NREM1-NREM4, movement time (M) and ? (unscored state). In this study, we considered W, REM and NREM 1, NREM 2 and SWS (NREM3 and NREM4 were considered as one sleep state) sleep states and segments belonging to other labels were rejected.

C. EEG data preparation

First, all EEG signals were detrended to remove linear trends from data. After data detrending, over-night EEG signals were divided into 30 seconds long epochs and normalized to zero mean and unit standard deviation so that within-subject variability, which is the results of subjects' personal physiological and psychological states, is disregarded. A rectangular window with length of 3000 samples (= 30 seconds) is used for extraction of EEG segments of each subject. We will call the obtained dataset as E with dimensions $m \times n$ where m refers to number of epochs extracted from EEG signals and n refers to the length of each epoch. Let $e_{ij}, i = 1, \dots, m, j = 1, \dots, n$ be the element of matrix E. At this step, the length of n is 3000 samples ($30s \times Fs = 3000$).

III. METHODOLOGY

The basic structure of the system proposed in this study is given in Fig. 1. It can be seen that it has three main modules: (1) signal denoising, (2) feature extraction/dimension reduction and (3) classification. In this section, we provide detailed explanation of each module.
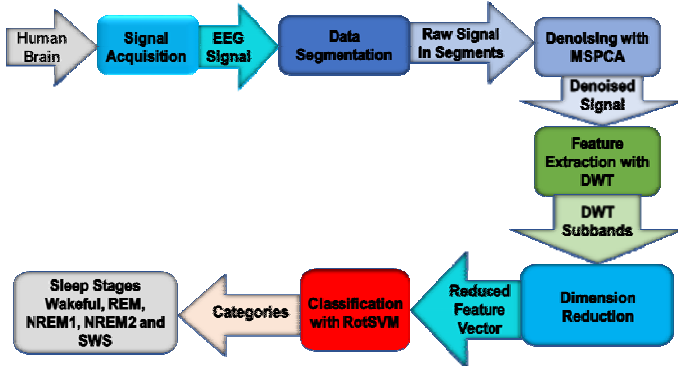

Fig. 1.Structural diagram of sleep stage classification system.

A. Module 1: Signal Denoising with MSPCA

The EEG signals are distorted by different types of noise and artefacts. General linear model to describe this signal in noise problem is:

$$E = S + N \qquad (1)$$

where S represent matrix containing clear source EEG signals and N represent noise and the aim is to remove N and obtain S which represents as close as possible undistorted brain waves. Generally, PCA can be used to solve signal in noise problem due to its ability to decorrelate correlated variables by extracting linear relationships among observations. PCA was already applied to determine sleep stage separation from EEG

signals in [38]. However, in order to perform these tasks in a better way for non-stationary signals whose behavior fluctuates over time and frequency, multiscale PCA (MSPCA) was proposed by Bakshi [11]. MSPCA was already successfully applied for denoising different biomedical signals, e.g. EEG [12], EMG [13] and ECG [14] where significant improvement in classification accuracy was achieved. MSPCA merges the capabilities of wavelet transform (WT) and PCA. The PCA extracts the relationship between multiple variables, whereas the WT decorrelate the auto-correlation among the measurements [39]. The steps of the MSPCA are illustrated in Fig. 2 and explained by the following algorithm [12] [40]:

1) Calculate wavelet decomposition at level J for each column in data matrix $\mathbf{X}$,
2) For $1 \leq m \leq J$, execute the PCA of the detail matrices $\mathbf{G}_m\mathbf{X}$ and choose a suitable number of significant principal components or reject the detail,
3) Execute the PCA of the approximation matrix $\mathbf{H}_J\mathbf{X}$ and choose a suitable number of principal components,
4) By inverting the wavelet transform $\mathbf{W}^T$, recover a new matrix from the reduced detail and approximation matrices,
5) Lastly, execute the PCA of that new matrix to form .

In MSPCA, wavelet analysis is used to decompose signal into different frequency bands (*here, we used 5 level decomposition*) and PCA is applied on the coefficients of generated bands and new coefficients are obtained.
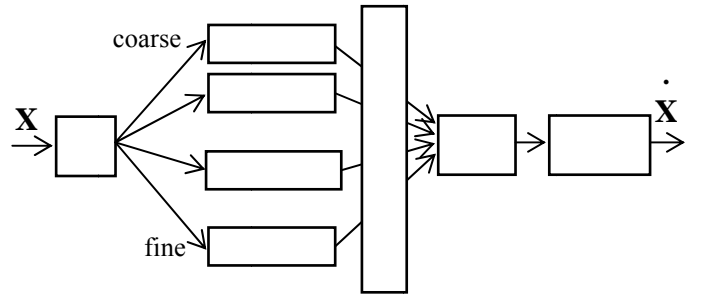

Fig. 2. Methodology for MSPCA *[12]*

B. Module 2: Feature Extraction with DWT

In this study, discrete wavelet transform (DWT) are used for feature extraction because of its capacity to give multi-resolution data representation. DWT decomposes signals into different frequency bands and extracts both time and frequency informative features from non-stationary signals, such as EEG. DWT first convolves vectors (EEG epochs) with low-pass filter (l) and with high-pass filter (h) and afterwards down-samples by using dyadic decimation and obtained vector are called approximation (cA_1) and detail (cD_1) respectively. In the next steps, DWT applies the same procedure on approximation vector (cA_1). DWT is widely used feature extraction technique which found application in wide area of researches and therefore it will not be discussed in detail. Instead, we forward readers interested in more details about DWT to Mallat's fast algorithm for DWT study [41].

C. Dimension Reduction

When selecting the basis ("mother") function which approximately matches the frequency characteristic of the EEG, i.e. delta, theta, alpha, beta, gamma, EEG signals were decomposed into six frequency bands in order to achieve required frequency resolution. Daubechies wavelet (db4) was selected as the basis function because of its recognized orthogonality property. In this study, six different statistical features were selected for EEG classification. The motivation to use signal statistics or to extract statistical features is to extract important information while reducing data dimensions. These statistical features are: a) Mean of coefficients' absolute values in each sub-band, b) Average power of the coefficients in each sub-band, c) Standard deviation of the coefficients in each sub-band, d) Ratio of absolute mean values of adjacent sub-bands, e) Skewness of every sub-band and f) Kurtosis of every sub-band.

Since we decomposed the EEG signal into six frequency bands we have cD_1, cD_2, cD_3, cD_4, cD_5, cD_6 and cA_6 from DWT decomposition. Hence, seven different features are extracted from each statistics (a), (b), (c), (e) and (f); six different features are extracted from (d) for each sub-band. Thus, in total, we have extracted 41 features for each epoch. In the present contribution, we also consider the ranking of these features to assess their relevance, we used information gain ranker method and found that all the features are relevant and important.

D. Module 3: Classification with RotSVM

After informative features are extracted, next step would be to feed these features into the classifier. SVM with RBF kernel was already applied on the same database in [9], but more complex feature extraction method was used, and the accuracy was low 89%. Therefore, we propose a new approach, i.e. modified SVM, which we can be called as rotational SVM (RotSVM). Motivation for this classifier was found in ensemble machine learning (EML) approach, referred to as a rotation forest, which was successfully applied in [42], and one can consider rotational SVM as type of rotation forest proposed by Rodriguez et.al. in 2006 [43]. In the present contribution, we have attempted to further improve state-of-the-art results reported in the literature. RotSVM can be understood as an EML method where SVM is trained on different feature subsets. Since SVM is one of the most celebrated and popular classification algorithm in the field of ML, we do not give details on the SVM classifier, but instead we forward the readers interested in knowing more about SVM to consult [44].

Here we present RotSVM algorithm. Let $f_i, i = 1, \ldots, m$ be the i-th row of matrix F and let $H = [h_1, \ldots, h_m]^T$ be vector with class labels for each row (epoch) where $h_i, i = 1, \ldots, m$ takes values from one of five class labels $SL = \{W, REM, NREM1, NREM2, SWS\}$. The classifier for each feature subset is the same, namely, SVM with polynomial kernel. Also, we denote the total number of SVM classifiers as X, which we train in parallel.

Training phase: For the x-th classifier, where $x = 1, \ldots, X$,

first decouple the matrix F into J subsets (submatrices) where each subset contains $n = 35 / J$ features. Let $F_j, j = 1, \ldots, J$ be j-th feature subset to train on x-th SVM classifier. For each F_j, arbitrarily choose a non-empty subset of classes, with size of $(3m) / 4$ to perform bagging and denote this subset as F_j'. In the next step, PCA should be applied on F_j' and let the new generated matrix be F_j''. In the subsequent step, store F_j'' in a sparse rotation matrix Ψ_x as:

$$\Psi_x = diag\left\{F_1'', \ldots, F_J''\right\} = \begin{bmatrix} F_1'' & [0] & \ldots & [0] \\ [0] & F_2'' & \ldots & [0] \\ \vdots & \vdots & \ddots & \vdots \\ [0] & [0] & \ldots & F_J'' \end{bmatrix} \quad (2)$$

In the next step, columns of Ψ_x matrix should be reordered to match the original features and let such matrix be Ψ_x'. In the last step, x-th classifier is built by employing $[F \cdot \Psi_x']$ and H as a training set. Repeat previous procedure for all SVMs.

Testing phase: In the classification phase, for a given epoch f, probability $p(x \cdot \Psi_x')$ is given by the x-th SVM classifier to the hypothesis that f belongs to θ_t, where t is one of the classes from SL, and the confidence for each class is computed by using the average combination method:

$$\lambda_t(f) = \frac{1}{X} \sum_{x=1}^{X} p(x \cdot \Psi_x'), t = SL(1), \ldots, SL(end) \quad (3)$$

Epoch f is associated with the class having the highest λ_t.

IV. EXPERIMENTAL RESULTS

A. Experimental Setup

In this section, we demonstrate how the proposed system can be used to solve the practical problem of sleep stage classification. To verify the performance of the proposed system, three different approaches were adopted. In the first approach, we evaluate the performances of the proposed system on each subject separately, where the system was trained and tested on the subject's own data and we refer to this approach as *Subject-Specific* approach (SSA). Records obtained from 20 subjects were employed and therefore 20 different datasets were generated. Two PSGs per subject were recorded for SC group, from where we randomly selected 10 subjects, and EEG data from these 2 PSGs were collected into one dataset per subject.

One may argue that this approach is biased since the system is trained and tested on data obtained from the same subject. Therefore, in our second approach, we attempt to avoid any possible such bias and therefore we adopted *Test-Group-Specific* classification approach (TGSA), where we generated two different datasets. The first dataset contained all epochs from the first study (SC) group and the second dataset contained all epochs from the second study (ST) group. In this approach, we aimed to see if the proposed system is robust to subjects with similar clinical history that is if the system is

robust when it is applied on any healthy subject or any subject with insomnia.

However, one may argue that this is approach is also biased since it was applied separately on each study group (subject with similar health conditions). Therefore, we have adopted the third approach which we call *Grand-Subjects-Specific* classification approach (GSSA).

A one-way analysis of variance (ANOVA) was carried out at 95% confidence level and p = 0.05 for ensuring statistical validation. Features are checked if p > 0.5 from the feature matrices and it is found p < 0.5 for 41 features. Moreover two-way-ANOVA test indicates that there was no benefit in adding more number of features while deriving discriminate feature vectors. Minimal drop of accuracy, significant at p = 0.05 was observed while decreasing feature dimension below 41. Therefore, significant features from where DWT sub-band features were evaluated were taken as 41.

In all three approaches, 10-fold CV was used, meaning that dataset was divided randomly into 10 different fold and 9 of these folds were used for training and remaining 10th fold was used for testing. This procedure was repeated 10 times, and at the end, average accuracy was computed.

In this study RotSVM, explained in Section III, was used for classification. SVM was trained using *sequential minimization algorithm (SMO)* [45] using polynomial kernel. To evaluate plausibility and efficiency of the system proposed system, performances were evaluated in terms sensitivity [46], Cohen's kappa coefficient κ [47] and overall accuracy. Sensitivity refers to fraction of positives that are correctly classified by the system. Cohen's kappa coefficient κ evaluates performance agreement between the proposed system and experts and gives more intuitive measure or the system overall performances (0-0.2: slight, 0.21-0.4: fair, 0.41-0.6: moderate, 0.61-0.8: substantial, 0.81-1: almost perfect agreement [48]). Accuracy is the number of epochs correctly classified by the system divided by the total number of epochs in dataset.

B. Performance Evaluation of Proposed System

Subject-Specific approach (SSA): Performance results for each subject are summarized in Figures 3, 4 and 5. Fig. 3 shows classification accuracy values while Fig. 4 shows sensitivity values for each of five different classes and average accuracy and sensitivity respectively for every subject. Fig. 5 shows Cohen's kappa coefficient κ values for each of 20 subjects. It can be seen From Fig. 3 that accuracy for detection of NREM1 sleep stages for six subjects was below 70 %, and this was the case when number of NREM1 epochs was evidently smaller when compared to number of epochs of the other sleep stages. Accuracy in detection of NREM1 sleep stages was above 80 % when the number of NREM1 sleep stages was similar to the number of NREM2 and SWS sleep stages. Fig.4 also confirms previous explanation and from this figure, it can be seen that sensitivity values were low for subjects with low accuracy rates, which is consistent with accuracy results. From Fig. 5, it can bee see that Cohen's kappa statistical values were above 0.8, from where it can be

seen that the proposed system is in almost perfect agreement with experts labeling.

Test-Group-Specific approach (TGSA): We were concerned that the aforementioned approach may be biased, since the proposed system was trained and tested on the data from the same subject. Therefore, we adopted the second approach to avoid such possible bias. Here, we generated two different datasets for two different studies explained in Section II. For that reason, we evaluated the system performance using data from all subjects from two different groups separately. 10-fold CV was adopted for training and testing. The obtained results for sensitivity and specificity are given in Table I. It is easy to see that the overall accuracy values are 95.89% for SC, and 93.37% for ST, indicating the high performances of the system. Cohen's coefficient value for the SC dataset is 0.92 and for the ST dataset is 0.91 indicating that this system is in almost perfect agreement with the experts.

Grand-Subjects-Specific approach (GSSA): Since one may also argue that the previous approach is also biased since it is evaluated in the subjects with similar health condition, we conducted the third experiment where the system was evaluated on all data collected from both groups. 10-fold CV was adopted for system training and testing. Obtained resulted are summarized in Table II. Sensitivity for this approach was 84.46% and the overall accuracy was 91.1%. Cohen's kappa coefficient was 0.88 showing also in this case almost perfect agreement with experts.
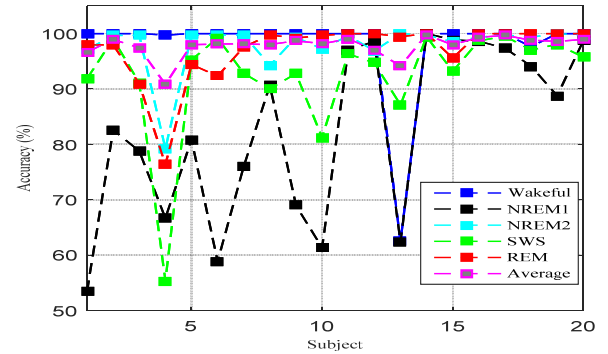

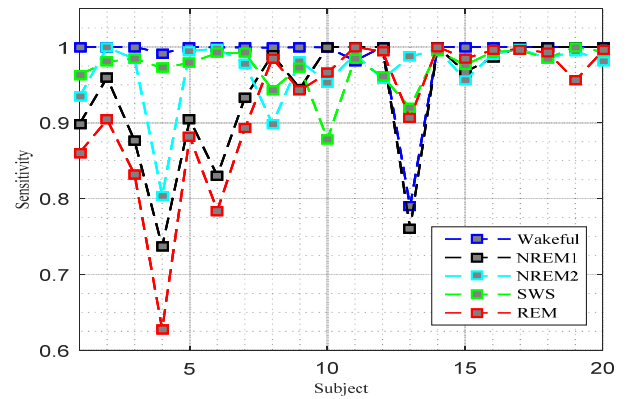Fig. 3 Accuracy values for SSA
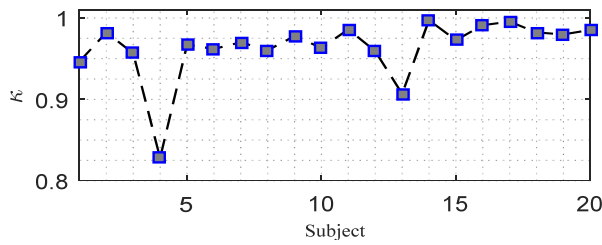

Fig. 4 Sensitivity values for SSA

Fig. 5. Cohen's kappa coefficient values for SSA

TABLE I
PERFORMANCE OF RotSVM FOR TGSA (%)

	Metric	W	N1	N2	SWS	REM	Avg.
S	Sensitivity	99.92	49.62	95.62	86.13	93.3	84.92
C	Accuracy	99.58	97.86	98.08	98.43	97.64	98.32
S	Sensitivity	96.54	79.55	96.97	89.14	94.7	91.32
T	Accuracy	99.43	97.05	96.72	96.08	96.99	97.25

W = wakeful, N1 = NREM1, N2 = NREM2, Avg. = Average

TABLE II
PERFORMANCE OF RotSVM FOR GSSA (%)

Metric	W	N1	N2	SWS	REM	Avg.
Sensitivity	98.59	48.78	95.54	97.94	94.59	84.46
Accuracy	99.24	95.23	95.81	95.91	95.11	96.26

TABLE III
CONFUSION MATRIX FOR TGSA (SC)

	W	NREM1	NREM2	SWS	REM
W	30421	19	0	0	6
NREM1	140	845	0	40	678
NREM2	15	7	9425	251	159
SWS	0	11	398	2738	32
REM	17	136	90	17	3622

TABLE IV
CONFUSION MATRIX FOR TGSA (ST)

	W	NREM1	NREM2	SWS	REM
W	920	20	0	2	11
NREM1	15	599	1	6	132
NREM2	0	0	3691	127	0
SWS	2	7	168	1650	24
REM	0	84	0	20	1860

TABLE V
CONFUSION MATRIX FOR GSSA

	W	NREM1	NREM2	SWS	REM
W	7178	93	0	0	10
NREM1	112	1198	1	87	1058
NREM2	15	8	13065	428	159
SWS	0	35	667	4265	63
REM	9	169	87	51	5530

TABLE VI
COMPARISON OF PERFORMANCE OF PREVIOUS STUDIES

Methods	Accuracy
Entropy metrics, J-means approach [8]	81%
Hybrid features, Artificial neural networks [49]	81.55%
Energy features, Recurrent neural classifier [50]	87.20%
Graph features, SVM [9]	88.90%
Spectral Features, Bootstrap aggregating [10]	86.53%
Temporal features and hierarchical decision tree [51]	77.98%
Fuzzy logic based iterative method [36]	74.50%
Multiscale entropy, LDA [22]	83.60%
Proposed Method	91.10%

C. Comparison with the existing methods

In this section, the classification performance of the proposed method is compared with some existing approaches. All these methods utilize the same database and are based on EEG signals. The overall classification accuracies of these methods are listed in Table VI. As it can be seen easily from the table, our method has better performance than the previous studies. Standard sleep stage classification methods proposed in literature generally need data from different biological signals (EEG, EMG, EOG, ECG, etc.) to extract informative features what usually lowers the sleep quality due to multiple electrodes which need to be attached to the body. Therefore, there is a need to single channel sleep monitoring. System proposed in this study requires only one EEG channel (Pz-Oz). In literature, few different systems based on single EEG channel were proposed. The proposed method has the best overall performance with the overall accuracy of 95.89% for *TGSA* (SC), 93.37% for *TGSA* (SC) and 91.1% for *GSSA*.

Flexer et.al. [52] used Gaussian observation hidden Markov model and reported accuracy of around 80 % for three-class classification (wakefulness, deep and REM sleep). Berthomier, et. al. [36] considered in their study PSG from 15 healthy subjects and proposed a system based on Automatic Sleep EEG Analysis (ASEEGA) and reported Cohen's kappa coefficient of 0.72 for five-class classification (W, REM, NREM1, NREM2 and SWS). Koley B. and Dey D. proposed a system based on SVM and recursive feature elimination for sleep stage classification and reported average kappa value of 0.8572. Liang et.al [22] in their study also proposed a system based on only one EEG channel which uses multiscale entropy and autoregressive model. They considered all-night PSG recordings from 20 healthy subject and reported average sensitivity of 0.836 and Cohen's kappa value of 0.81 for five-state classification (W, REM, NREM1, NREM2 and SWS) which are considerably lower when compared to results obtained in this study. Zhu, et.al. [9] , considered in their study 8 PSG records from Sleep-EDF database and proposed a system based on difference visibility graph and SVM to classify the sleep stages and reported the accuracy and Cohen's kappa coefficient of 87.5 % and 0.81. Although these results yielded almost equally good results (accuracy of 87.5 % vs. 90.18 %, Cohens' kappa of 0.81 vs 0.86), the number of subjects considered in [9] varied considerably (8 vs. 20).

From Tables I-V, it can be seen that sensitivity value for detection of NREM1 sleep stage are considerably lower when compared with detection of other sleep stages. This is mainly due to the fact that NREM1 and REM exhibit similar EEG patterns since NREM1 is transition stage between wakefulness and different sleep stages just like REM is transitional between sleeps stages and wakefulness state (backward direction) [4]. Therefore, low sensitivity and accuracy values were obtained for NREM1. Systems which require multiple channels also experience the same problem what can be seen from similar studies reported in literature [9, 27].

D. Discussion

This study proposes a robust system for automatic classification of five different sleeps stages, namely

wakefulness, NREM1, NREM2, SWS and REM, using single-channel EEG signals. The obtained results, explained in Section III.C demonstrate that the correct classification of five sleep stages is possible using only Pz-Oz EEG channel. As a matter of fact, along with high accuracy values obtained in all three approaches, high sensitivity values were also obtained. Cohen's kappa coefficients were higher than 0.8 indicating perfect agreement with experts sleep stage labeling. Besides, the proposed method contains two wavelet analyses, one in the MSPCA for denoising and one for feature extraction. The MSPCA combines the features of wavelet analysis and PCA through decomposing every variable on a sym4 mother wavelet. Also, DWT is used for feature extraction with db4 mother wavelet. It should also be noted that only first four moments were extracted from EEG signal frequency sub-bands and these features are sufficient to extract the most information from EEG data. Since MCPCA is used for denoising, the proposed system is robust for noise. Furthermore, SVM is a robust classifier and by using ensemble of SVM which is RotSVM, the classifier will be more robust. These can be seen from the performance of the proposed method.

As a classifier, rotational SVM, which can be thought as an ensemble of SVM, is proposed in the present contribution. Although it is not mentioned here, standard SVM with the same features as input and same parameter setup was also evaluated in this study and considerable improvements in performances were noticed when rotational SVM was used as classifier comparted to traditional SVM approach. This finding may reflect the relevance and the success of the operation of a RotSVM scheme over traditional SVM in such a way that RotSVM keeps the information relevant for sleep-stage classification, so that the correct decision on classification can be made. Thus, a picture that emerges is that RotSVM can be applicable for the sleep-stage classification tasks.

One may argue that PCA is not relevant for classification tasks since discriminatory information is not involved in computation of the optimal axis rotation. Therefore, numerous linear transformation substitutes to PCA established on discrimination conditions were proposed in literature [53]. Also, PCA may impose additional problems in classification due to data dimensionality reduction, if used, since some of the components which are neglected due to small variance can play important role in classification task. However, in this study, we kept all principal components in classification module. Experimental results showed improvement in performances when this approach was adopted. Although, not reported in this study, we also tried to use several other methods in lieu of PCA (random projection, normalization, wavelets), but the highest performances were obtained when PCA was used.

This study proposes an automated system for sleep stage classification. It proposes a new approach for standard SVM classifier, which can be called as rotational SVM. Classification results in Figures 3-5 and Tables I-V demonstrate system's high performances. Even though,

performances were lower when the system was trained and tested on data multiple subjects (Test-Group-Specific approach and Grand-Subjects-Specific approach), result were still satisfactory. Remarkably, results obtained in this study demonstrate that only one single EEG channel can be used for sleep stage classification. For that reason, proposed system can be considered as promising tool for sleep monitoring.

E. Advantages and Disadvantages of the Proposed System

The core advantage of the proposed system over other approaches found in the existing literature is in its ability to incorporate all the available information from only one EEG channel to make an accurate decision on the sleep stages. It does this by extracting the features of the highest relevance from each time-frequency band and feeding them into the efficient classifier. The second advantage lies in its ability not to use only the subject's own data, but the model can be trained on any subject; and still be efficiently applied only on the small portion of the new test data from new subject, what saves recording time, and thus, clinicians' time. The third is the high classification accuracy rates and performance results. The disadvantage of the proposed system is that our proposed model is non-linear, what might introduce additional computational complexity. In our future work, we will focus on the design of the linear model for accurate sleep stage classification.

V. CONCLUSION

In the present study, automated system for sleep stage classification system using only one channel EEG is proposed. The proposed system consists of three modules, preprocessing (denoising), feature extraction (using DWT) and classification using rotational SVM (ensemble machine learning tool). Overall classification accuracy, average sensitivity and Cohen's kappa coefficient obtained with this system were 91.1%, 84.46 % and 0.88 respectively for classification of five different sleeps stages (wakefulness, NREM1, NREM2, SWS and REM). Furthermore, system proposed in this study requires single EEG channel signal which further simplifies sleep stage monitoring. Since manual sleep stage classification is often time-consuming and subjective, and therefore prone to errors, system proposed in this study can be considered as tool in clinical and clinical and home-care application to discriminate specific patterns such as fatigue, drowsiness and/or various sleep disorders (e.g., sleep apnea) in near real-time. As a conclusion, the system proposed in this study has the potential to substantially enhance sleep monitoring systems.

VI. REFERENCES

[1] T. J. Sejnowski and A. Destexhe, "Why do we sleep?," Brain Research, vol. 886, p. 208–223, 2000.

[2] R. Stickgold, "Sleep-dependent memory consolidation," Nature, vol. 437, p. 1272–1278, 2005.

[3] A. Rechtschaffen and A. Kales, A manual of standardized terminology, techniques, and scoring system for sleep stages of human subjects, Los Angeles, CA: UCLA, Brain Research Institute/Brain Information Service, 1968.

[4] S.-L. Himanena and J. Hasan, "Limitations of Rechtschaffen and Kales," Sleep Medicine Reviews, vol. 4, no. 2, p. 149–167, 2004.

[5] C. Iber, S. Ancoli-Israel, A. L. Chesson and S. F. Quan, The AASM Manual for Scoring of Sleep and Associated Events-Rules: Terminology and Technical Specification, Darien, IL, USA: American Academy of Sleep Medicine , 2007.

[6] U. R. Acharya , S. Bhat, O. Faust, H. Adeli, E. C. P. Chua, W. J. E. Lim and J. E. W. Koh, "Nonlinear Dynamics Measures for Automated EEG-Based Sleep Stage Detection," European Neurology, vol. 74, pp. 268-287, 2015.

[7] L. Fraiwan, K. Lweesy, N. Khasawneh, H. Wenz and H. Dickhaus, "Automated sleep stage identification system based on time–frequency analysis of a single EEG channel and random forest classifier," Computer Methods and Programs in Biomedicine, vol. 108, no. 1, pp. 10-19, 2012.

[8] J. L. Rodríguez-Sotelo, A. Osorio-Forero, A. Jiménez-Rodríguez, D. Cuesta-Frau, E. Cirugeda-Roldán and D. Peluffo, "Automatic sleep stages classification using EEG entropy features and unsupervised pattern analysis techniques," Entropy, vol. 16, no. 12, p. 6573–6589, 2014.

[9] G. Zhu, Y. Li and P. P. Wen, "Analysis and Classification of Sleep Stages Based on Difference Visibility Graphs From a Single-Channel EEG Signal," IEEE Journal of Biomedical and Health Informatics, vol. 18, no. 6, pp. 1813 - 1821 , 2014.

[10] A. R. Hassan and A. Subasi, "A decision support system for automated identification of sleep stages from single-channel EEG signals," Knowledge-Based Systems, vol. 128, pp. 115-124, 2017.

[11] B. R. Bakshi, "Multiscale PCA with application to multivariate statistical process monitoring," AIChE Journal, vol. 44, no. 7, p. 1596–1610, 1998.

[12] J. Kevric and A. Subasi, "The Effect of Multiscale PCA De-noising in Epileptic Seizure Detection," Journal of Medical Systems, vol. 38, no. 10, p. 131, 2014.

[13] E. Gokgoz and A. Subasi, "Effect of multiscale PCA de-noising on EMG signal classification for diagnosis of neuromuscular disorders," Journal of Medical Systems , vol. 38, no. 4, p. 31, 2014.

[14] E. Alickovic and A. Subasi, "Effect of Multiscale PCA De-noising in ECG Beat Classification for Diagnosis of Cardiovascular Diseases," Circuits, Systems, and Signal Processing, vol. 34, no. 2, p. February, 2015.

[15] S. Gunes, K. Polat, M. Dursun and S. Yosunkaya, "Examining the relevance with sleep stages of time domain features of EEG, EOG, and chin EMG signals," in 14th National Biomedical Engineering Meeting, 2009. BIYOMUT 2009, Balcova, Izmir, 2009.

[16] Y. Li, K. M. Wong and H. d. Bruin, "Electroencephalogram signals classification for sleep-state decision – a Riemannian geometry approach," IET Signal Processing, vol. 6, no. 4, p. 288–299, 2011.

[17] V. Bajaj and R. B. Pachori, "Automatic classification of sleep stages based on the time-frequency image of EEG signals," Computer Methods and Programs in Biomedicine, vol. 112, no. 3, p. 320–328, 2013.

[18] X. Long, P. Fonseca, J. Foussier, R. Haakma and R. M. Aarts, "Sleep and Wake Classification With Actigraphy and Respiratory Effort Using Dynamic Warping," IEEE Journal of Biomedical and Health Informatics, vol. 18, no. 4, pp. 1272 - 1284 , 2014.

[19] R. B. Duckrow and H. P. Zaveri, "Coherence of the electroencephalogram during the first sleep cycle," Clinical Neurophysiology, vol. 116, no. 5, p. 1088–1095, 2005.

[20] C. E. Rasmussen and C. K. I. Williams, Gaussian Processes for Machine Learning, Cambridge, MA: The MIT Press, , 2006.

[21] U. R. Acharya, E. C. Chua, K. C. Chua, L. C. Min and T. Tamura, "Analysis and automatic identification of sleep stages using higher order spectra," International Journal of Neural Systems , vol. 20, no. 6, pp. 509-521, 2010.

[22] S.-F. Liang, C.-E. Kuo, Y.-H. Hu, Y.-H. Pan and S.-F. Liang , "Automatic Stage Scoring of Single-Channel Sleep EEG by Using Multiscale Entropy and Autoregressive Models," IEEE Transactions on Instrumentation and Measurement, vol. 61, no. 6, pp. 1649 - 1657 , 2012.

[23] T. Shimada, T. Shiina and Y. Saito, "Detection of characteristic waves of sleep EEG by neural network analysis," IEEE Transactions on Biomedical Engineering, vol. 47, no. 3, pp. 369 - 379, 2000.

[24] M. E. Tagluk, N. Sezgin and M. Akin, "Estimation of Sleep Stages by an Artificial Neural Network Employing EEG, EMG and EOG," Journal of Medical Systems, vol. 34, no. 4, pp. 717-725, 2010.

[25] H.-t. Hu, R. Talmon and Y.-L. Lo, "Assess Sleep Stage by Modern Signal Processing Techniques," IEEE Transactions on Biomedical Engineering, vol. 62, no. 4, pp. 1159 - 1168, 2015.

[26] T. Willemen, D. Van Deun, V. Verhaert, M. Vandekerckhove, V. Exadaktylos, J. Verbraecken, S. Van Huffel, B. Haex and J. V. Sloten, "An Evaluation of Cardiorespiratory and Movement Features With Respect to Sleep-Stage Classification," IEEE Journal of Biomedical and Health Informatics, vol. 18, no. 2, pp. 661 - 669 , 2014.

[27] S. Enshaeifar, S. Kouchaki, C. C. Took and S. Sanei, "Quaternion Singular Spectrum Analysis of Electroencephalogram with Application in Sleep Analysis," IEEE Transactions on Neural Systems and Rehabilitation Engineering, pp. 11-10, 2015.

[28] L. G. Doroshenkov, V. A. Konyshev and S. V. Selishchev, "Classification of human sleep stages based on EEG processing using hidden Markov models," Biomedical Engineering, vol. 41, no. 1, pp. 25-28, 2007.

[29] H. G. Jo, J. Y. Park, C. K. Lee, S. K. An and S. K. Yoo, "Genetic fuzzy classifier for sleep stage identification," Computers in Biology and Medicine, vol. 40, no. 7, p. 629–634, 2010.

[30] C. M. Held, J. E. Heiss, P. A. Estevez, C. A. Perez, M. Garrido, C. Algarín and P. Peirano, "Extracting Fuzzy Rules From Polysomnographic Recordings for Infant Sleep Classification," IEEE Transactions on Biomedical Engineering , vol. 53, no. 10, pp. 1954-1962, 2006.

[31] L. Breiman, "Random Forest," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.

[32] B. Kemp, A. H. Zwinderman, B. Tuk, H. A. Kamphuisen and 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, no. 9, pp. 1185-1194 , 2000.

[33] B. Kemp, "The Sleep-EDF Database [Expended]," [Online]. Available: http://www.physionet.org/physiobank/database/sleep-edfx/. [Accessed 1 DEcember 2015].

[34] PhysioNet, "PhysioBank Archive Index," 1 December 2015. [Online]. Available: http://www.physionet.org/physiobank/database/.

[35] 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 and H. E. Stanley , "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals," Circulation, vol. 101, no. 23, pp. e215-e220 , 2000.

[36] C. Berthomier, X. Drouot, M. Herman-Stoica, P. Berthomier, J. Prado, D. Bokar-Thire, O. Benoit, J. Mattout and M.-P. d'Ortho, "Automatic Analysis of Single-Channel Sleep EEG: Validation in Healthy Individuals," Sleep, vol. 30, no. 11, p. 1587–1595, 2007.

[37] B. van Sweden, B. Kemp, H. A. Kamphuisen and E. A. Van der velde, "Alternative electrode placement in (automatic) sleep scoring (Fpz-Cz/Pz-Oz versus C4-A1/C3-A2)," Sleep, vol. 13, no. 3, pp. 279-283, 1990.

[38] C. Vural and M. Yildiz, "Determination of Sleep Stage Separation Ability of Features Extracted from EEG Signals Using Principle Component Analysis," Journal of Medical Systems , vol. 34, pp. 83-89, 2010.

[39] B. R. Bakshi, "Multiscale PCA with Application to Multivariate Statistical Process Monitoring," AlChE, vol. 44, no. 7, pp. 1596-

1610, 1998.

[40] M. Aminghafari, N. Cheze and J.-M. Poggi, "Multivariate denoising using wavelets and principal component analysis," Computational Statistics & Data Analysis, vol. 50, pp. 2381-2398, 2006.

[41] S. Mallat, A Wavelet Tour of Signal Processing, Academic Press, 2008.

[42] E. Alickovic and A. Subasi, "Breast cancer diagnosis using GA feature selection and Rotation Forest," Neural Computing & Applications , 18 November 2015.

[43] J. J. Rodriguez, L. I. Kuncheva and C. J. Alonso, "Rotation Forest: A New Classifier Ensemble Method," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 10, pp. 1619 - 1630 , 2006.

[44] C. M. Bishop, Pattern Recognition and Machine Learning, Springer Science+Business Media, LLC, 2006.

[45] J. C. Platt, "Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines," in Advances in Kernel Methods – Support Vector Learning, B. Scholkopf, C. J. C. Burges and A. J. Smola, Eds., MIT Press, Platt, p. 185–208.

[46] D. G. Altman , Practical statistics for medical research, London, U.K.: Chapman & Hall, 1991.

[47] J. Cohen, "A Coefficient of Agreement for Nominal Scales," Educational and Psychological Measurement, vol. 20, no. 1, pp. 37-46 , 1960 .

[48] J. R. Landis and G. G. Koch, "The Measurement of Observer Agreement for Categorical Data," Biometrics, vol. 33, no. 1, pp. 159-174, 1977.

[49] M. Ronzhina, O. Janoušek, J. Kolářová, M. Nováková, P. Honzík and I. Provazník, "Sleep scoring using artificial neural networks," Sleep Med Rev, vol. 16, no. 3, p. 251–263, 2012.

[50] Y.-L. Hsu, Y.-T. Yang , J.-S. Wang and C.-Y. Hsu , "Automatic sleep stage recurrent neural classifier using energy features of EEG signals," Neurocomputing, vol. 104, p. 105–114, 2013.

[51] S.-F. Liang, C.-E. Kuo , Y.-H. Hu and Y.-S. Cheng , "A rule-based automatic sleep staging method," J. Neurosci. Methods, vol. 205, no. 1, p. 169–176, 2012.

[52] 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 .

[53] Y. Chien and F. King-Sun, "On the generalized Karhunen-Loeve expansion," IEEE Transactions on Information Theory, vol. 13, no. 3, pp. 518 - 520, 1967.