Kernel Dimensionality Reduction on Sleep Stage Classification using ECG Signal

Sani M. Isa[1], Ito Wasito[2], Aniati Murni Arymurthy[3]

[1] Faculty of Computer Science, University of Indonesia
Depok 16424/Jawa Barat, Indonesia

[2] Faculty of Computer Science, University of Indonesia
Depok 16424/Jawa Barat, Indonesia

[3] Faculty of Computer Science, University of Indonesia
Depok 16424/Jawa Barat, Indonesia

Abstract

The purpose of this study is to apply Kernel Dimensionality Reduction (KDR) to classify sleep stage from electrocardiogram (ECG) signal. KDR is supervised dimensionality reduction method that retains statistical relationship between input variables and target class. KDR was chosen to reduce dimensionality of features extracted from ECG signal because this method doesn't need special assumptions regarding the conditional distribution, the marginal distribution, or both. In this study we extract 9 time and frequency domain heart rate variability (HRV) features from ECG signal of Polysomnographic Database from Physionet. To evaluate KDR performance, we perform sleep stage classification using kNN, Random Forest and SVM method, and then compare the classification performance before and after dimensionality reduction using KDR. Experimental result suggested KDR implementation on sleep stage classification using SVM could reduce dimensionality of feature vector into 2 without affecting the classification performance. KDR performance on Random Forest and k Nearest Neighbour classification only show slight advantage compared to without implementing KDR.

Keywords: *dimensionality reduction, KDR, polysomnography, ECG, sleep stage.*

1. Introduction

Sleep is a behavioral state that is a natural part of every individual's life. Human spend about one-third of their lives asleep. Sleep is not just something to fill time when a person is inactive. Sleep is important for normal motor and cognitive function. Sleep actually appears to be required for survival [1]. Sleepiness problem may be associated with concentration difficulty, memory lapses, loss of energy, fatigue, lethargy, and emotional instability. The prevalence of problem sleepiness is high and has serious consequences, such as drowsy driving or workplace accidents and errors. Of the more than 70 known sleep disorders, the most common are obstructive sleep apnea, insomnia, narcolepsy, and restless legs syndrome. Large numbers of individuals suffering from these sleep disorders are unaware and have not been diagnosed or treated for their disorder [2].

The most widely used monitoring technique for assessing suspected sleep disorders is polysomnography (PSG) performed in a sleep laboratory. PSG includes the three measures used to assess the sleep state and to determine the sleep stage: electroencephalography (EEG), submental electromyography (EMG), and electrooculography (EOG). During a typical PSG, respiratory effort, airflow at the nose and mouth, oxyhemoglobin saturation, electrocardiogram (ECG), and leg movements are also assessed, with continuous recording throughout the night [3]. PSG analysis produces a summary of the patient's sleep architecture.

The identification of macrostructure and microstructure of sleep are still relying on visual scoring that requires long and accurate work by specialized personnel. Therefore, some algorithms and procedures were developed to perform automatic sleep scoring. Further, many efforts were dedicated to search of signals that can be reliably recorded through wearable devices. Thus, everyone can do sleep monitoring in their home with ease.

ECG recording is one of the simple and efficient technologies in sleep disorders detection. Cyclic variations in RR intervals of ECG signals have been reported to be associated with sleep apnea and different sleep stages. Various studies have confirmed that several new methods could possibly recognize sleep apnea and sleep stages from

heart rate variability (HRV) [4]. In 2010, Yilmaz proposed the use of three features derived from the RR-interval, i.e. median, inter-quartil range (IQR), and mean absolute deviation (MAD) for sleep apnea and sleep stage classification [5]. In the same year, Bsoul proposed client-server architecture to determine sleep efficiency based on the sleep stage classification using ECG signal [6]. Those studies have shown that the standard ECG recording can provide comparable results to standard PSG analysis.

Latest technological advance on health monitoring device has enabled us to have personal diagnostic tools such as portable polysomnography or portable electrocardiograph device. Some of this device even has intelligent features so that it can be used as simple diagnostic tools. Such intelligent features would require data and algorithm optimization because the limitation of portable devices on processing data. Dimensionality reduction is one of technique to optimize data size. After dimensionality reduction the data size will reduce to smaller dimension without changing characteristic of data. In this study we use KDR method to reduce dimensionality of feature extracted from ECG. KDR is supervised dimensionality reduction method that retains statistical relationship between input variables and target class. KDR was chosen to reduce dimensionality of features extracted from ECG signal because this method doesn't need special assumptions regarding the conditional distribution.

The objective of this study is to assess the performance of KDR method on sleep stage classification using ECG signal. KDR was used to reduce dimensionality of feature vector. Performance assessment of KDR method was performed by comparing the performance of classification without dimensionality reduction and with dimensionality reduction. We use four time domain and five frequency domain HRV features in classification.

2. Methodology

The schematic diagram of the system, which is used in this study is shown in Fig. 1. There are five main stages on this system, preprocessing, feature extraction, dimensionality reduction, classification, and performance evaluation. Detail explanation of each stage will be covered in next part of this paper.

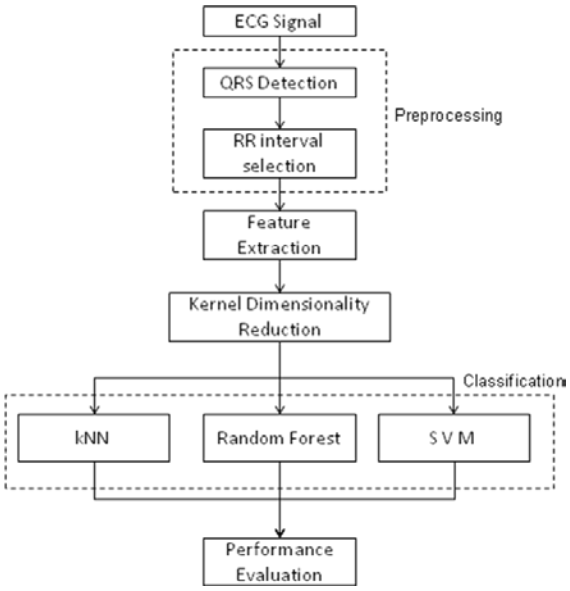

Fig. 1 Schematic diagram of the system used in this study.

2.1 Subjects

Data used in the experiment is MIT-BIH Polysomnographic Database. The database contains over 80 hours' worth of four-, six-, and seven-channel polysomnographic recordings from 16 subjects, each with an ECG signal annotated beat-by-beat, and EEG and respiration signals annotated with respect to sleep stages and apnea for every 30s epoch [7]. In this database, all 16 subjects were male, aged 32 to 56 (mean age 43), with weights ranging from 89 to 152 kg (mean weight 119 kg). Records slp01a and slp01b are segments of one subject's polysomnogram, separated by a gap of about one hour; records slp02a and slp02b are segments of another subject's polysomnogram, separated by a ten-minute gap. The remaining 14 records are all from different subjects.

The reference annotations were made by human experts on the basis of simultaneously recorded signals. All recordings include an ECG signal, an invasive blood pressure signal, an EEG signal, and a respiration signal. The six- and seven-channel recordings also include a respiratory effort signal derived by inductance plethysmography; some include an EOG signal and an EMG signal and the remainder includes a cardiac stroke volume signal and an earlobe oximeter signal. The sleep stage distributions over the total 10087 epochs were: W=2951, NREM 1=1750, NREM 2=3939, NREM 3=499, NREM4=222, REM=726. Epoch with MT (movement time) annotation was not included in this study.

2.2 Preprocessing

All the features used in this study are based on QRS detection times. A 'QRS detection time' is the time of occurrence of the QRS complex in an ECG signal. In this study, QRS detection times were generated automatically for all recordings using Engelse and Zeelenberg algorithm [8]. This algorithm provides detection times that occur close to the onset of the QRS complex. RR-intervals defined as the interval from the peak of one QRS complex to the peak of the next. RR-interval sequences generated from QRS detection times could contained physiologically unreasonable times because signal quality limitation which leads to errors in the automatically generated QRS detections. In this study only RR intervals between 0.5 and 1.5 were processed on the stage of feature extraction [5].

2.3 Feature Extraction

The system used in this study is an epoch-based system that processes features based on the timing of QRS complexes. For each epoch with duration of 30 seconds, we extract HRV features. HRV measures can be divided into two broad categories: time domain measures and frequency domain measures [9]. The commonly time domain heart rate variability statistics are defined in Table 1. Commonly used frequency domain measures are defined in Table 2 [10]. Due to short epoch duration, only short-term time-domain and frequency domain HRV features are used in this study (AVNN, SDNN, rMSSD, pNN50, TOTPWR, VLF, LF, HF, LF/HF).

Table 1: Commonly used time-domain measures

Measurements	***Explanation***
AVNN*	Average of all RR intervals
SDNN*	Standard deviation of all RR intervals
SDANN	Standard deviation of the averages of RR intervals in all 5-minute segments of a 24-hour recording
SDNNIDX	Mean of the standard deviations of RR intervals in all 5-minute segments of a 24-hour recording
rMSSD*	Square root of the mean of the squares of differences between adjacent RR intervals
pNN50*	Percentage of differences between adjacent RR intervals that are greater than 50 ms; a member of the larger pNNx family

* Short-term HRV statistics

Traditionally frequency domain measures are calculated by resampling the original RR interval series and then applying the fast Fourier transform or autoregressive spectral estimation (the maximum entropy method). This resampling can cause attenuation in the high frequency components [10]. To eliminate the need for evenly sampled data required by Fourier or maximum entropy methods, we use the Lomb periodogram to calculate frequency domain spectra for unevenly sampled data [11]. Due to the low frequency resolution of a 30 seconds power spectral estimation, RR-interval spectrum calculated based on 5 epochs centered on the epoch of interest. This approach reduces the time-localization of the sleep stage information, but increased the spectral resolution [12].

Table 2: Commonly used frequency-domain measures

Measurements	***Explanation***
TOTPWR*	Total spectral power of all RR intervals up to 0.04 Hz
ULF	Total spectral power of all RR intervals up to 0.003 Hz
VLF*	Total spectral power of all RR intervals between 0.003 and 0.04 Hz
LF*	Total spectral power of all RR intervals between 0.04 and 0.15 Hz.
HF*	Total spectral power of all RR intervals between 0.15 and 0.4 Hz
LF/HF*	Ratio of low to high frequency power

* Short-term HRV statistics

To remove subject dependency, the RR interval series need to be normalized. A normalized RR interval series were calculated by dividing by the mean of RR interval, which produce RR interval with unity mean. Only time domain HRV features were use the normalized version of RR interval series. Normalization for frequency domain HRV features was achieved by dividing VLF, LF, and HF by the TOTPWR.

2.4 Kernel Dimensionality Reduction

In supervised dimensionality reduction, we assume data consists of *(X,Y)* pairs, where *X* is a *n x m*-dimensional explanatory variable and *Y* is an *l*-dimensional response. The variable *Y* may be either continuous or discrete. Solution to the feature selection problem will be a linear combination of *X*. We assume that there is an *r*-dimensional subspace $S \subset R^m$ such that the following equality holds for all x and y:

$$P_{Y|X}(y,x) = p_{Y|\Pi_S X}(y \mid \Pi_S x), \quad (1)$$

where $\prod S$ is the orthogonal projection of *R* onto *S*. The subspace *S* is called the effective subspace for regression/classification. Based on observations of *(X,Y)* pairs, we wish to recover a matrix whose columns span *S*. KDR approach the problem using semiparametric

statistical framework—therefore no assumptions regarding the conditional distribution of $p_{Y|\prod_S X}(y|\prod_S x)$ or the distribution of $p_X(x)$ of X.

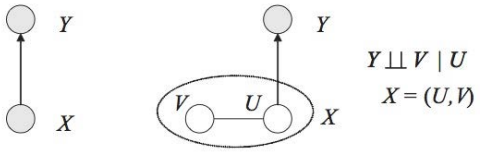

Fig. 2 Illustration of dimensionality reduction for regression/classification.

The notion of effective subspace can be formulated in terms of conditional independence. Let $Q = (B,C)$ be an m-dimensional orthogonal matrix such that the column vectors of B span the subspace S (thus B is $m \times r$ and C is $m \times (m - r)$), and define $U = B^T X$ and $V = C^T X$. Because Q is an orthogonal matrix, we have $p_X(x) = p_{U,V}(u,v)$ and $p_{X,Y}(x,y) = p_{U,V,Y}(u,v,y)$. Eq. (1) is equivalent to

$$p_{Y|U,V}(y|u,v) = p_{Y|U}(y|u). \tag{2}$$

Eq. (2) shows that the effective subspace S causes Y and V conditionally independent for given U (see Fig. 2). Mutual information provides another perspective of equivalence between conditional independence and effective subspace. It is well known that

$$I(Y,X) = I(Y,U) + E_U\big[I(Y|U,V|U)\big], \tag{3}$$

where $I(Y,X)$ is the mutual information between X and Y. Because Eq. (1) implies $I(Y,X) = I(Y,U)$, the effective subspace S is characterized as the subspace which retains the entire mutual information between X and Y, or equivalently, such that $I(Y|U, V|U) = 0$. This lead to the conditional independence of Y and V given U.

KDR use cross-covariance operators on RKHSs to describe conditional independence of random variables. Let (H,k) reproducing kernel Hilbert space of functions on a set Ω with positive definite kernel $k : \Omega \times \Omega \to R$ and an inner product $\langle .,. \rangle_H$. The most significant characteristic of a RHKS is the reproducing property:

$$\langle f, k(\cdot, x)\rangle_{\mathcal{H}} = f(x) \quad \text{for all } x \in \Omega \text{ and } f \in \mathcal{H}. \tag{4}$$

Fukumizu et al. uses the Gaussian kernel $k(x1, x2) = \exp(-\|x1 - x2\|^2/2\sigma^2)$ [13]. They also show that for probability-determining kernel spaces, the effective subspace S can be characterized in terms of the solution to the following minimization problem:

$$\min_S \Sigma_{YY|U}, \quad \text{subject to} \quad U = \Pi_S X. \tag{5}$$

Centralized Gram matrix is then used for estimation of the operator:

$$\begin{aligned}\hat{K}_Y &= \big(I_n - \tfrac{1}{n}\mathbf{1}_n\mathbf{1}_n^T\big)G_Y\big(I_n - \tfrac{1}{n}\mathbf{1}_n\mathbf{1}_n^T\big),\\ \hat{K}_U &= \big(I_n - \tfrac{1}{n}\mathbf{1}_n\mathbf{1}_n^T\big)G_U\big(I_n - \tfrac{1}{n}\mathbf{1}_n\mathbf{1}_n^T\big)\end{aligned} \tag{6}$$

where $1_n = (1,\ldots,1)^T$, $(G_Y)_{ij} = k_1(Y_i, Y_j)$ is the Gram matrix of the samples of Y, and $(G_U)_{ij} = k_2(U_i, U_j)$ is given by the projection $Ui = B^T X_i$. With a regularization constant $\varepsilon > 0$, the empirical conditional covariance matrix $\Sigma_{YY|U}$ is then defined by

$$\begin{aligned}\hat{\Sigma}_{YY|U} &:= \hat{\Sigma}_{YY} - \hat{\Sigma}_{YU}\hat{\Sigma}_{UU}^{-1}\hat{\Sigma}_{UY}\\ &= (\hat{K}_Y + \varepsilon I_n)^2 - \hat{K}_Y\hat{K}_U(\hat{K}_U + \varepsilon I_n)^{-2}\hat{K}_U\hat{K}_Y.\end{aligned} \tag{7}$$

The size of $\Sigma_{YY|U}$ in the ordered set of positive definite matrices can be evaluated by its determinant. Using the Schur decomposition,

$$\det(A - BC^{-1}B^T) = \det\begin{pmatrix} A & B \\ B^T & C \end{pmatrix}/\det C, \tag{8}$$

we have

$$\det\hat{\Sigma}_{YY|U} = \det\hat{\Sigma}_{[YU][YU]}/\det\hat{\Sigma}_{UU}, \tag{9}$$

where $\Sigma_{[YU][YU]}$ is defined by

$$\begin{aligned}\hat{\Sigma}_{[YU][YU]} &= \begin{pmatrix} \hat{\Sigma}_{YY} & \hat{\Sigma}_{YU} \\ \hat{\Sigma}_{UY} & \hat{\Sigma}_{UU} \end{pmatrix}\\ &= \begin{pmatrix} (\hat{K}_Y + \varepsilon I_n)^2 & \hat{K}_Y\hat{K}_U \\ \hat{K}_U\hat{K}_Y & (\hat{K}_U + \varepsilon I_n)^2 \end{pmatrix}.\end{aligned} \tag{10}$$

The objective function then symmetrize by dividing by the constant $\det\Sigma_{YY}$, which give

$$\min_{B\in\mathbb{R}^{m\times r}} \frac{\det\hat{\Sigma}_{[YU][YU]}}{\det\hat{\Sigma}_{YY}\det\hat{\Sigma}_{UU}}, \quad \text{where } U = B^T X. \tag{11}$$

This minimization problem referred with respect to the choice of subspace S or matrix B as Kernel Dimensionality Reduction (KDR).

To determine the effective dimensions, in this study the original feature vector dimension was reduced to 2, 3, and 4 dimension using KDR. The results of classification using new reduced feature vector then compared with the result of classification using original feature vector.

2.5 Classification

We applied three classification methods, which were kNN as the baseline method, random forest and SVM. The last two methods were selected because they have different performance on imbalanced dataset. As we know, sleep stage data has characteristics of imbalanced dataset due to imbalanced distribution of each sleep stage. Table 3 shows sleep stage distribution for each record. From table 3, we can see that most of records dominated by sleep stage NREM 2.

To train the classifier, 66 % of the epochs were randomly selected and the remaining epochs presented to the testing phase. The training data was chosen in such a way that the distribution of sleep stages on each record retained. Cross validation with 10 fold was then applied to assess how the predictive model will generalize to an independent data set. We performed 10 repetitions in the testing phase to ensure that the results of classification describe the actual condition.

Table 3: Sleep distribution for each record

Records	***Sleep Stage Distribution (%)***					
	Wake	***NREM 1***	***NREM 2***	***NREM 3***	***NREM 4***	***REM***
slp01a	2.9	0.4	43.9	19.7	27.6	5.4
slp01b	50.6	7.6	34.7	-	-	7.1
slp02a	12.6	5.2	58.0	1.4	0.6	22.1
slp02b	39.6	5.3	44.2	-	-	10.9
slp03	18.1	15.2	44.8	11.3	-	10.7
slp04	22.2	8.2	61.7	4.6	-	3.2
slp14	45.1	26.2	17.7	4.2	1.7	5.1
slp16	45.4	15.6	26.2	3.2	0.3	9.4
slp32	61.3	4.2	25.0	6.8	2.7	-
slp37	10.8	3.0	84.6	-	-	1.6
slp41	29.0	29.6	28.1	1.7	-	11.6
slp45	0.9	0.9	62.3	8.9	12.6	14.4
slp48	27.9	31.8	35.9	0.3	-	4.1
slp59	30.7	23.0	21.1	11.0	6.6	7.7
slp60	40.7	47.8	7.1	-	-	4.5
slp61	17.3	12.3	45.4	14.3	-	10.7
slp66	39.4	32.8	26.6	1.1	-	-
slp67x	46.7	26.0	26.7	0.7	-	-
Overall	30.07	16.39	38.56	5.37	1.32	6.07

Classification carried out on each epoch to determine the sleep stage of those epochs. Two set of classification output was used to differentiate six sleep stages (Wake, NREM 1, NREM 2, NREM 3, NREM 4, and REM) and four sleep stages (Wake, Light Sleep, Deep Sleep, and REM). Light Sleep stage is the combination of NREM 1 and NREM 2 sleep stage, whereas Deep Sleep stage is the combination of NREM 3 and NREM 4 sleep stage. To find out the subject dependency factor, the classification was performed using two scenarios, which are subject specific and subject independent classification. On subject specific classification, data used for training and testing phase obtained from the same record, whereas on subject independent classification, data obtained from the combination of all records.

2.6 Performance Evaluation

We apply two classification performance measurements, which are accuracy and Cohen's Kappa coefficient. Accuracy is used as a statistical measure of how well a classification test correctly identifies or excludes a condition. Cohen's kappa coefficient is a statistical measure of inter-rater agreement or inter-annotator agreement[13] for qualitative (categorical) items. It is generally thought to be a more robust measure than simple percent agreement calculation since κ takes into account the agreement occurring by chance.

Cohen's kappa measures the agreement between two raters who each classify N items into C mutually exclusive categories. The equation for κ is:

$$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)}, \tag{12}$$

where *Pr(a)* is the relative observed agreement among raters, and *Pr(e)* is the hypothetical probability of chance agreement, using the observed data to calculate the probabilities of each observer randomly saying each category. If the raters are in complete agreement then $\kappa = 1$. If there is no agreement among the raters then $1 > \kappa \geq 0$. Table 4 shows kappa coefficient interpretation [13].

Table 4: Kappa coefficient interpretation

Kappa	***Agreement***
< 0	Less than chance agreement
0.0 – 0.20	Slight agreement
0.21 – 0.40	Fair agreement
0.41 – 0.60	Moderate agreement
0.61 – 0.80	Substantial agreement
0.81 – 1.00	Almost perfect agreement

3. Results and Discussion

3.1 Cross Validation

Table 5 shows 10 fold cross validation result. In this study we considered overall classification accuracy. Cross

validation results suggest that classification on subject specific data gives better classification accuracy than subject independent data. This results indicate that classification performance depend on subjects. Overall accuracy of classification using 4 output classes (Wake, Light Sleep, Deep Sleep, and REM) is better than 6 output class (Wake, NREM1, NREM2, NREM3, NREM4, REM).

Table 5: Cross validation result with k=10

Dimension Reduction	*Effective Dimension*	*Subject*	*Output Class*	*Accuracy (Stdev)*		
				kNN	*Random Forest*	*SVM*
None	9	Independent	6	40.1 (4.26)	46.55 (4.39)	44.13 (3.35)
			4	53.42 (4.14)	60.5 (4)	57.07 (2.22)
		Specific	6	57.9 (6.06)	64.35 (5.22)	57.32 (3.91)
			4	68.17 (5.86)	74.33 (4.86)	66.37 (4.43)
KDR	2	Independent	6	37.34 (4.18)	39.52 (4.11)	46.11 (3.61)
			4	51.66 (3.91)	55.16 (4.33)	61.21 (3.35)
		Specific	6	59.24 (6.64)	60.96 (6.38)	64.26 (5.27)
			4	68.4 (6.27)	70.1 (6.19)	72.83 (5.1)
	3	Independent	6	39.59 (4.78)	42.41 (4.47)	44.22 (3.78)
			4	53.88 (4.63)	56.92 (4.23)	59.14 (3.13)
		Specific	6	60.34 (5.79)	60.86 (5.88)	59.55 (4.42)
			4	70.93 (5.64)	71.75 (5.58)	69.79 (5.34)
	4	Independent	6	40.72 (4.77)	44.06 (4.32)	44.4 (3.8)
			4	53.79 (4.28)	58.76 (3.98)	59.06 (2.94)
		Specific	6	59.57 (6.4)	61.06 (5.97)	56.25 (3.49)
			4	69.75 (5.6)	71.1 (5.95)	66.48 (4.14)

3.2 Split/Train Test

Table 6 and Table 7 show overall classification accuracy and kappa statistics of split train/test respectively. Overall classification accuracy of split train/test shows consistent result compared to cross validation result. Almost on all experimental setup, Kappa statistics of split train/test on subject specific data shows fair to moderate agreement (0.21 – 0.48) except on SVM classification using effective dimension = 4 and output class = 6 which gives value of Kappa statistics = 0.14. KDR implementation on SVM classification shows significant Kappa statistics gain, whereas on other classification method show only slight gain.

According to Table 6, we cannot determine which method gives the best overall classification accuracy because some of values from one method are overlapped with the other methods. To determine the best overall classification accuracy we employ head to head comparison between each method. We count correct classification result from each experiment then apply normalization so that the value in range 0 to 100. This value is then used as a basis for classification performance comparation.

Table 6: Classification accuracy of split train/test

Dimension Reduction	*Effective Dimension*	*Subject*	*Output Class*	*Accuracy (Stdev)*		
				kNN	*Random Forest*	*SVM*
None	9	Independent	6	39.53 (2.22)	45.54 (2.34)	42.96 (1.39)
			4	52.59 (2.06)	60.31 (1.99)	56.85 (0.76)
		Specific	6	57.84 (2.74)	64.08 (3.13)	57.35 (1.92)
			4	68.31 (2.57)	73.43 (2.56)	66.3 (1.65)
KDR	2	Independent	6	37.37 (1.92)	39.05 (2)	45.19 (1.94)
			4	51.96 (2.11)	54.6 (2.31)	60.53 (1.52)
		Specific	6	58.95 (2.93)	60.28 (2.98)	63.05 (2.56)
			4	68.05 (2.83)	69.65 (2.95)	71.81 (2.33)
	3	Independent	6	39.18 (2.15)	41.97 (2.2)	43.8 (1.52)
			4	53.01 (2.03)	56.72 (2.5)	58.56 (1.46)
		Specific	6	59.59 (3.03)	59.85 (3.23)	58.1 (2.19)
			4	69.87 (2.67)	71.18 (2.64)	68.14 (2.12)
	4	Independent	6	39.88 (2.24)	43.36 (2.04)	44.26 (1.66)
			4	53.12 (2.02)	57.74 (2.19)	58.35 (1.25)
		Specific	6	58.58 (2.97)	60.47 (3.26)	55.07 (1.46)
			4	69.22 (2.63)	70.37 (2.75)	65.4 (1.87)

Table 7: Kappa statistics of split train/test

Dimension	*Effective Dimensio*	*Subject*	*Output*	*Kappa (Stdev)*

Reduction	*n*		*Class*	*kNN*	*Random Forest*	*SVM*
None	9	Independent	6	0.16 (0.03)	0.21 (0.04)	0.09 (0.02)
			4	0.19 (0.03)	0.26 (0.04)	0.05 (0.02)
		Specific	6	0.33 (0.04)	0.41 (0.05)	0.21 (0.03)
			4	0.4 (0.05)	0.48 (0.05)	0.22 (0.03)
KDR	2	Independent	6	0.13 (0.03)	0.14 (0.03)	0.17 (0.03)
			4	0.18 (0.03)	0.2 (0.04)	0.21 (0.03)
		Specific	6	0.36 (0.04)	0.37 (0.05)	0.36 (0.05)
			4	0.4 (0.05)	0.43 (0.05)	0.42 (0.05)
	3	Independent	6	0.16 (0.03)	0.17 (0.03)	0.12 (0.02)
			4	0.2 (0.03)	0.23 (0.04)	0.14 (0.03)
		Specific	6	0.37 (0.05)	0.36 (0.05)	0.22 (0.04)
			4	0.43 (0.05)	0.45 (0.05)	0.29 (0.05)
	4	Independent	6	0.17 (0.03)	0.19 (0.03)	0.12 (0.03)
			4	0.2 (0.03)	0.24 (0.04)	0.13 (0.03)
		Specific	6	0.35 (0.05)	0.36 (0.05)	0.14 (0.03)
			4	0.42 (0.05)	0.43 (0.05)	0.21 (0.04)

3.3 Head To Head Comparisons

Table 8 shows head to head comparison result between kNN versus Random Forest, kNN versus SVM, and Random Forest versus SVM. Table 7 suggested that Random Forest and SVM outperform kNN on all experiment setup. Classification using Random forest shows better result than SVM on classification without dimension reduction. On the contrary, SVM gives better result than Random Forest after dimension reduction using KDR with effective dimension 2 and 3. From table 8 we can see that Random Forest optimal on original data without dimension reduction, whereas SVM achieved best result on reduced dimension data with effective dimension = 2.

Table 8: Head to head comparison between kNN, Random Forest and SVM

Dim Reduction	*Effective Dim*	*Subject*	*Output Class*	*kNN vs RF*		*kNN vs SVM*		*RF vs SVM*	
None	9	Inde-pendent	6	1	99	16	84	73	27
			4	0	100	5	95	87	13
		Specific	6	7	93	52	48	88	12
			4	6	94	62	38	88	12
		Average		3	97	34	66	84	16
KDR	2	Inde-pendent	6	25	75	0	100	1	99
			4	19	81	0	100	0	100
		Specific	6	28	72	14	86	22	78
			4	28	72	13	87	22	78
		Average		25	75	7	93	11	89
	3	Inde-pendent	6	12	88	1	99	21	79
			4	9	91	2	98	24	76
		Specific	6	38	62	59	41	58	42
			4	32	68	56	44	71	29
		Average		23	77	30	70	44	56
	4	Inde-pendent	6	6	94	2	98	29	71
			4	8	92	0	100	42	58
		Specific	6	25	75	69	31	80	20
			4	31	69	68	32	81	19
		Average		18	82	35	65	58	42

3.4 KDR Performance

Fig. 2 to Fig. 5 shows chart of KDR performance on subject independent and subject specific data. KDR implementation on kNN and Random Forest classification didn't show overall accuracy gain because classification performance was decreased after KDR implemented, conversely KDR implementation on SVM classification shows overall accuracy gain. The best overall classification accuracy when KDR implemented was achieved by SVM method with effective dimension = 2. This result suggested that KDR implementation on sleep stage classification using SVM have successfully increased overall classification accuracy.

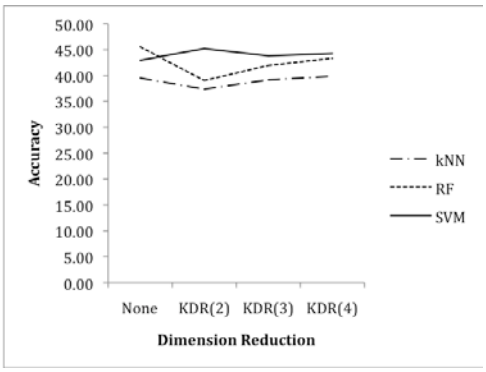

Fig. 2 KDR performance on subject independent data with output class = 6.

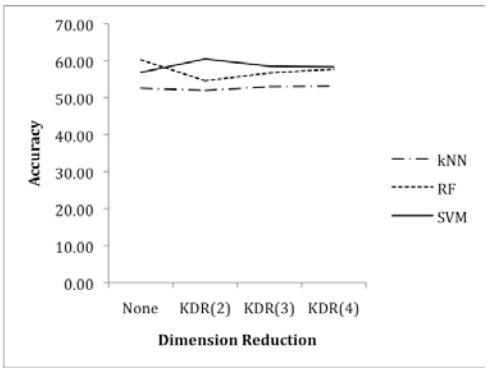

Fig. 3 KDR performance on subject independent data with output class = 4.

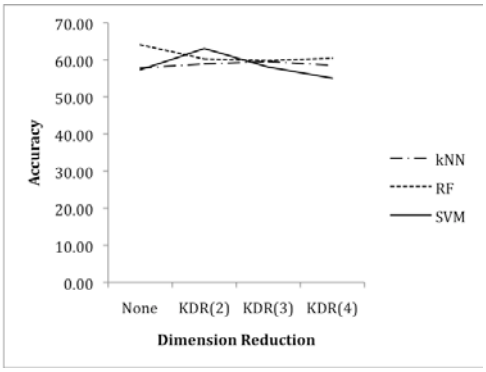

Fig. 4 KDR performance on subject specific data with output class = 6.

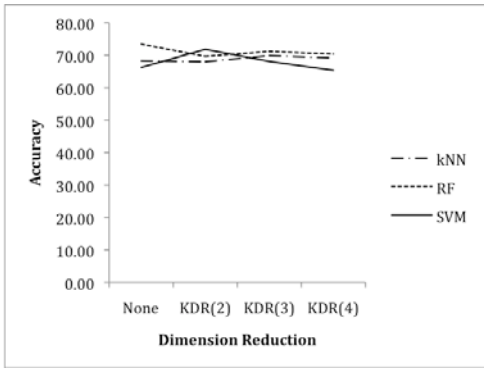

Fig. 5 KDR performance on subject specific data with output class = 4.

4. Conclusions

We have presented KDR implementation on sleep stage classification using ECG signal. From the experimental result, we have shown that KDR have successfully reduce feature vector dimension of data while maintain overall classification accuracy. This indicates KDR maintain relation between input and output variables while reducing dimension. The most optimal result of KDR implementation achieved by SVM classification with effective dimension = 2. These results demonstrate considerable potential in applying KDR in sleep stage classification using ECG signal.

In this study we also have shown that the head to head comparison could be used as comparation method for special case i.e. small classification performance differences between two methods. In the future research, we will apply KDR in other problems such as sleep apnea detection. Next research will also study the influence of EDR (ECG derived respiratory) signal on sleep stage classification or sleep apnea detection.

References

[1] Sleep Disorder Overview. [http://www.neurologychannel.com/sleepdisorders/index.shtml]
[2] Sleep Apnea: What Is Sleep Apnea? [http://www.nhlbi.nih.gov/health/dci/Diseases/SleepApnea/SleepApnea _WhatIs.html]
[3] Armon C, Johnson G.K, Roy A, Nowack W.J: Polysomnograph. [http://emedicine.medscape.com/article/1188764-overview]
[4] T. Penzel, J. McNames, P. de Chazal, B. Raymond, A. Murray, G. Moody, “Systematic comparison of different algorithms for apnoea detection based on

electrocardiogram recordings", Medical and Biological Engineering and Computing 2002, 40:402–407.
[5] Yilmaz B, Asyali MH, Arikan E, Yetkin S, Ozgen F, "Sleep stage and obstructive apneaic epoch classification using single-lead ECG", Biomed Eng Online 2010, 9(1):39.
[6] Bsoul M, Minn H, Nourani M, Gupta G, Tamil L., "Real-time sleep quality assessment using single-lead ECG and multi-stage SVM classifier", in Conf Proc IEEE Eng Med Biol Soc, 2010, 1:1178-1181.
[7] Y Ichimaru, GB Moody, "Development of the polysomnographic database on C D-ROM. Psychiatry and Clinical Neurosciences 1999", 53:175-177.
[8] Chazal P, Penzel T, and Heneghan C, "Automated detection of obstructive sleep apnoea at different time scales using the electrocardiogram", Physiological Measurement. 2004, 25: 967-983.
[9] American Academy of Sleep Medicine Task Force, "Sleep-related breathing disorders in adults: Recommendations for syndrome definition and measurement techniques in clinical research Sleep", 1999, 22:667–689.
[10] Heart Rate Variability Analysis with the HRV Toolkit: Basic Time and Frequency Domain Measures. [http://www.physionet.org/tutorials/hrv-toolkit/]
[11] G.D. Clifford, Azuaze F, McSharry P, "Advanced Methods And Tools for ECG Data Analysis", Artech House Publishers, Norwood; 2006.
[12] S. Redmond, C Heneghan, "Electrocardiogram-BasedAutomatic Sleep Staging in Sleep Disordered Breathing", Computers in Cardiology 2003, 30:609-612.
[13] K. Fukumizu, F.R. Bach, a nd M.I. Jordan, "Kernel Dimensionality Reduction for Supervised Learning", in Proc. NIPS, 2003.
[14] Lewicke A, Sazonov E, Corwin MJ, Neuman M, Schuckers, "Sleep versus wake classification from heart rate variability using computational intelligence: consideration of rejection in classification models", IEEE Trans Biomed Eng 2008, 55(1):108-118.

Sani M. Isa hold Bachelor degree in Mathematics from Faculty of Natural Science, Padjadjaran University, Indonesia and M aster degree from Faculty of Computer Science, University of Indonesia.

Ito Wasito hold Ph.D in Computer Science, School of Computer Science and Information Systems, Birkbeck College, University of London, United Kingdom.

Aniati Murni Arymurthy hold Bachelor degree in Electrical Engineering from Faculty of Engineering, University of Indonesia, Master degree in Computer and I nformation Science, The Ohio State University, United States of America, and Ph.D in Optoelectronics and Laser Application, University of Indonesia.