Hybrid Dimensionality Reduction of Multi-sets Using Nature Inspired Algorithms and Discriminant Canonical Correlation Analysis for Automatic Sleep Stage Classification

Pimporn Moeynoi[1]*   **Yuttana Kitjaidure[1]**

[1]Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand
* Corresponding author's Email: 56601036@kmitl.ac.th

Abstract: Automatic sleep stage classification using various signals is an important tool to assess sleep disorders and sleep quality. However, the numerous variables from multi sources lead to the classification problems. To improve the problem, this paper proposes a new hybrid dimensionality reduction by combining the feature selection based on the nature-inspired algorithms (NIAs) and multi-sets transformation technique by Canonical Correlation Analysis (CCA)/ Discriminant Canonical Correlation analysis (DCCA). The NIAs is first adopted to generate the updated population positions in each dataset, and then the CCA/DCCA is used to fuse the selected subsets. The proposed algorithm performance is demonstrated on sleep-wake detection and multi-class sleep stage classification. Furthermore, the proposed method called Dissimilarity Binary Grey Wolf Optimization – Discriminant Canonical Correlation Analysis (DisBGWO-DCCA), modified by using the differential evolution (DE) technique based on the similarity of Jaccard coefficient, provides the best classification accuracies with 97.75% of the sleep-wake detection and 95.85% of the multi-class sleep stage classification when comparing with other single dimensionality reduction approaches[1-2]. Moreover, the proposed algorithm achieves the computational cost better than the conventional NIAs as shown later in experiment. Our experiment is also operated on both healthy subjects and sleep disorder patients with efficient sleep stage classification.

Keywords: Hybrid dimensionality reduction, Binary grey wolf optimization, Discriminant canonical correlation analysis, Dissimilarity binary grey wolf optimization - Discriminant canonical correlation analysis, Sleep stage classification.

1. Introduction

The identification of autonomic sleep pattern has become an interesting research topic because sleep stage scoring is an essential process to estimate sleep disorders and sleep quality. In clinical studies, sleep stages can be categorized into the waking, rapid eye movement (REM), and non-rapid eye movement (NREM) stages. These are commonly conducted by a gold standard sleep recording called polysomnographic recording (PSG). The PSG is obtained from various signals such as electroencephalography (EEG), electrooculography (EOG), electrocardiography (ECG), blood oxygen saturation level, the respiratory signal, and other

signals that relate to fluctuations in systems during sleep. Generally, the sleep stage scoring is manually scored by sleep experts based on two standards of manual visual sleep scoring including the Rechtschaffen and Kales (R&K) and the American Academy of Sleep Medicine (AASM), which needs hours and many signals for sleep stage identification. For this reason, the development of an automated sleep stage system is a demand to assist with diagnosis. The automatic system would need to be easily applied in clinical practice or sleep studies, but high performance with less computing time is also a necessity to increase the quality of diagnoses.

Several studies attempted to develop automatic sleep scoring systems based on multi-sources. For example, a few of works employed the feature

extraction techniques that included frequency domain and time-frequency domain [1], and non-linear technique [2, 3] to classify multi class sleep stage. These techniques achieved with the accuracies ranging from 85% to 90% that still was not satisfied. The work of [4] utilized the EEG, EOG and EMG signals and employed the wrapper techniques of sequential forward selection (SFS) and sequential backward selection (SBS) to improve the sleep stage classification performance. The similar study in [5] proposed the combination of filtering and wrapper to find the candidate feature subset. The experiment results showed that these methods achieved high performance on sleep and wake stages, but they gave a low accuracy on multi-class sleep stage classification. The main drawback of these methods is the obtained features ignored the relation between features that has importance for analysis in multi different sources. To solve the problem, the obtained feature sets are combined by the linear correlation technique as proposed in the efficient study of [6]. Although the classical correlation technique has the ability to solve the issue of relationship between features, this technique still has a disadvantage that may cause some information to be lost in the high dimensional data. Another drawback of using wrapper method for the feature selection is that the process is required the time-consuming computation. To improve these problems, we propose the new hybrid dimensionality reduction algorithm that will be later described.

From above mentions, the multiple signals are required for the sleep stage classification processing. Multisource signals reveal that the numerous variables derived from several sources contain different functional properties and provide specific information about the objects. However, too many variables lead to the curse of dimensionality problems in automated sleep stage classification such as expensive computation and decreasing accuracy. In order to enhance these problems, the automatic systems need the dimensionality reduction to deal with high dimension data. The dimensionality reduction methods can be divided into two approaches, which are single and hybrid approaches. The first one, the single approach performs either feature transformation or feature selection to be a low space. The other one, the hybrid approach reduces dimension by combining more than one approach. Recently, the hybrid dimensionality reduction has been popularly proposed to improve the classification efficiency in the fields of data medical [7, 8].

The feature selection can be divided into three methods: filter, wrapper, and embedded. This paper only focuses on the wrapper method using heuristic

search algorithms because these algorithms offer local optimum results which can achieve good results and feasible computation. The natural inspired algorithms (NIAs) – the part of the meta-heuristic search algorithms including Binary Particle Swarm Optimization (BPSO) [9], Binary Cuckoo Search (BCS) [10], Artificial bee colony [11], and Binary Grey wolf Algorithm Optimization (BGWO) [12, 13], have become popular choices for feature selection that is able to search the optimal solutions and select the best subset, also explore the space efficiently. These techniques were compared and suggested in previous studies [12, 13] that the advantages of BGWO were easy to implement and high efficiency. Moreover, it has been the most recent NIAs techniques. The BGWO technique mimics the hunting behavior in the nature of a group of grey wolves similar to the classical grey wolf algorithm optimization technique. However, in the classical GWO, the solution is considered and updated in a continuous version while the solution of BGWO is considered in the discrete version. The binary values are required in feature selection therefore the update occurs at the corner of a hypercube. The BGWO for feature selection tasks has been revealed to achieve higher classification accuracy than BCO, BGA, and BPSO.

The feature transformation or feature extraction transforms high dimension space into the lower space by using the combination of the datasets. Generally, many works in [14 - 16] suggested that this method can improve the performance of machine learning system. For instance, canonical correlation analysis (CCA) technique aims to fuse multiple relevant data sets and find the maximal correlation between data sets. Although CCA has been used successfully in various machine learning applications, it still has limitations in multi-model classification performance. The works in [17 - 19] reported that the CCA did not utilize the information of the sample classes, so it affected the classification applications. To solve this problem, the study of [20] proposed the DCCA to combine the extracted features utilized the class information. The DCCA obtains the linear canonical correlation by finding the maximum correlation within the class and minimizing the correlation between the different classes. The discriminative power of this method is higher than that of the conventional CCA method proposed in [21 - 23] studies such as classification, regression, and clustering.

The conventional techniques for solving the dimensionality reduction of sleep stage classification have drawbacks in the field of multi-dataset. In
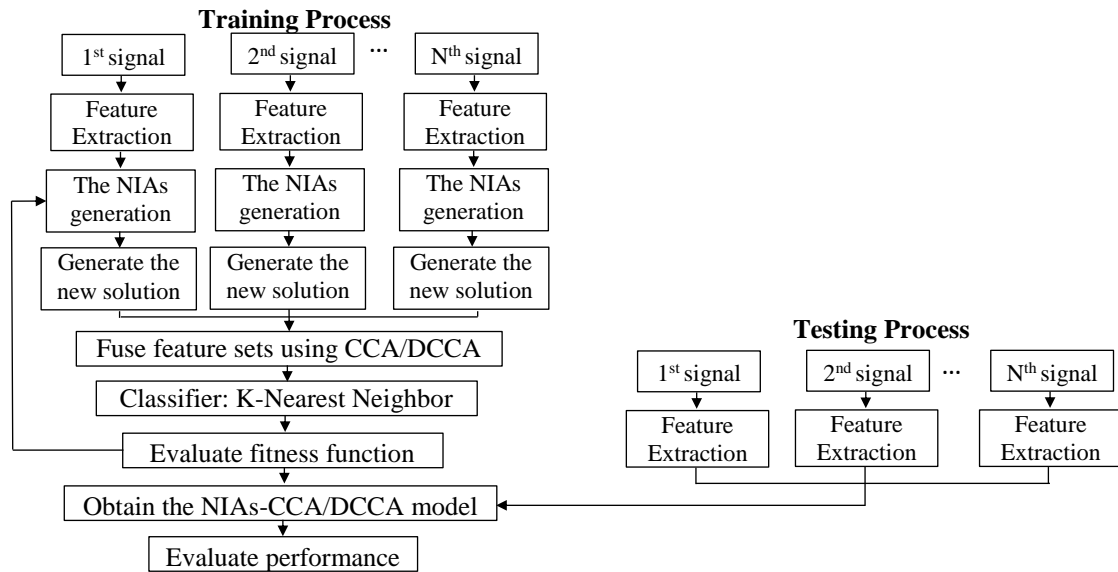
279

Training Process

Figure. 1 Block diagram of the proposed algorithm

feature selection process, the obtained features are considered from the relevance of each original feature with their output labels. Thus, the selected features ignore the correlation between features from different sources leading to poor generation in the test process. In addition, the feature transformation method is reported that an original information may be lost after transformation. To solve these problems, we propose the novel hybrid dimensionality reduction technique that integrates some of the NIAs feature selection techniques and the DCCA/CCA multi-sets transformation. This algorithm aims to achieve high classified accuracy and having minimal features, which can handle with the problem of time-consuming. Thus, the proposed algorithm attempts to compensate for the disadvantage in feature selection and feature transformation methods. This algorithm not only searches local features that allow better local discrimination, but also considers relations between feature sets that exist significantly in original information. This hybrid dimensionality reduction technique includes two main steps. The first step, the binary techniques of NIAs, including BPSO, BCS, EBCS, BGWO and modified BGWO, are developed to select the best set of feature in each dataset. In the second step, the CCA/DCCA is utilized to fuse the feature sets, then the fused vector is sent to evaluate by the classifier. The procedure of the proposed algorithm is shown in Fig. 1. We evaluate the efficiency of these algorithms in experiments of sleep wake detection and multiclass sleep stage classification. The comparative results show that the DisBGWO-DCCA, modified by using the differential evolution (DE) technique based on the dissimilarity of Jaccard coefficient achieves the best performance.

The rest of this paper is organized as follows. Section 2 reviews the background of the feature selection and muli-sets transformation techniques. Section 3 presents the proposed algorithm. Section 4 describes the experimental procedures. Section 5 shows the results of experiments. Finally, Section 6 concludes this study.

2. Background

2.1 The nature inspired algorithms for feature selection

This section reviews the popular and proficient NIAs for feature selection including BPSO, BCS, EBCS, and BGWO.

2.1.1. Binary particle swarm optimization (BPSO)

In feature selection, BPSO is more suitable technique than continuous PSO, which particles are updated by the best previous local position and the best global position. The study of [9] introduced that the elements in the position were either 1 or 0 indicated by the probability of the element in the position vector. The position (x_d) and velocity (V_d) of each particle (d^{th}) are updated as follows:

$$x_d(t+1) = \begin{cases} 1, if\ \dfrac{1}{1+e^{-V_d(t+1)}} \ge rand \\ 0, otherwise \end{cases} \quad (1)$$

$$V_d(t+1) = w \times V_d(t) + C_1 r_1 \left(pbest - x_d(t)\right) \\ + C_2 r_2 \left(gbest - x_d(t)\right) \quad (2)$$

, where t is t^{th} iteration, w is the weight constant, C_1 and C_2 are acceleration constants, r_1, r_2 and $rand$ are random values in [0,1], and $pbest$ and $gbest$ are the best previous position of particles and the best global position.

2.1.2. Binary Cuckoo search (BCS)

The study of [10] proposed the binary cuckoo search that utilized a technique called the levy flights random walk to explore the search space. The new solutions of BCS are updated and generated by binary values based on a sigmoid function as:

$$x(t+1) = \begin{cases} 1, if \ \dfrac{1}{1+e^{-x_d(t)}} \ge rand \\ 0, otherwise \end{cases} \quad (3)$$

$$x_d(t) = x_d(t-1) + \alpha \oplus Levy(\lambda) \quad (4)$$

, where x_d represents the value of each dth egg on the nest at iteration t, $rand$ is a random value in [0 1], $\alpha > 0$ is the step size scaling weight, and $\oplus$ is entry wise multiplication. $Levy$ is described as

$$Levy \sim u = t^{(-\lambda)}, 1 < \lambda \le 3 \quad (5)$$

, where u is step size and λ is Levy flight distribution. In the last step of each iteration, the nest with the worst quality eggs are replaced to new ones with a probability $P_a \in [0 \ 1]$.

2.1.3. Extended binary Cuckoo search (EBCS)

The classical BCS utilizes the random walk search technique that suffers low convergence because it jumps from one region to another one easily. To modify and improve BCS, the work in [24] added a local mutation procedure to extend the classical CS called EBCS. In each iteration of EBCS, a predefined number of mutation solutions (N_{mut}) are randomly selected as the input of the operation. Then, each element of solution is randomly selected and switched their binary values using the sigmoid function in Eq.(3). If the element value is 1, it will be converted to 0 and vice versa.

2.1.4. Binary grey wolf optimization (BGWO)

The GWO mimics the hunting mechanism of grey wolves in nature. The algorithm of classical GWO is described with the following steps; firstly, the GWO parameters such as the number of variables or dimensions and maximum number of iteration are initialized. Secondly, wolf solutions are randomly generated and evaluated with the number of the packs and the fitness value, respectively. Next, the three best wolves in the pack are identified as X_α, X_β and X_δ, and then the location of the wolves is updated using

$$X(t+1) = \frac{X_1(t) + X_2(t) + X_3(t)}{3} \quad (6)$$

The $X_1(t)$, $X_2(t)$, and $X_3(t)$ are defined as follows;

$$X_1(t) = X_\alpha(t) - A_1(C_1 X_\alpha(t) - X(t)) \quad (7)$$

$$X_2(t) = X_\beta(t) - A_2(C_2 X_\beta(t) - X(t)) \quad (8)$$

$$X_3(t) = X_\delta(t) - A_3(C_3 X_\delta(t) - X(t)) \quad (9)$$

The $A = 2ar_1 - a$ and $C = 2r_2$, where r_1, r_2 are random vectors in the range of [0, 1] and α is linearly decreased from 2 to 0. Finally, the fitness value is obtained and the best global values, X_α, X_β and X_δ, are updated. The BGWO procedure slightly differs from the classical GWO. This technique aims to search and identify the best subset of features in forms of binary values. Therefore, the updated binary vector is given:

$$x_d(t+1) = \begin{cases} 1, if \ \dfrac{1}{1+e^{-m\left(\left(\frac{X_1+X_2+X_3}{3}\right)-n\right)}} \ge rand \\ 0, otherwise \end{cases} \quad (10)$$

, where $rand \in [0 \ 1]$, m and n are the parameters of the sigmoid function, and $x_d(t+1)$ represents the generated binary value in dimension d^{th} at iteration $t+1$.

2.2 Jaccard dissimilarity coefficient

The binary dissimilarity measure has been proposed in the feature selection method for improving the problem of classification described in [11]. The study in [25] explained the similarity of two binary vectors (S_u and S_v) as

$$Similarity(S_u, S_v) = \frac{J_{11}}{J_{11} + J_{10} + J_{01}} \quad (11)$$

The J_{11} means the number of bits where $S_u = S_v = 1$, J_{10} means the number of bits where $S_u = 1$ and $S_v = 0$ and J_{01} means the number of bits where $S_u = 0$ and $S_v = 1$. Likewise, the dissimilarity (Dis) measure between S_u and S_v vector is defined by;

$$Dis\left(S_u, S_v\right) = 1 - \frac{J_{11}}{J_{11} + J_{10} + J_{01}} \quad (12)$$

2.3 CCA/DCCA for multi-sets transformation

The transformation using CCA/DCCA aims to combine relevant information from multiple vectors into a single vector. The CCA proposed by [28] used linear transformation to find the maximum correlation between two sets of variables. Suppose that a set of pair-wise samples $D = \{(X_{i1}, X_{i2})^n_{i=1}\}$, where $X_{i1} \in R^p$ and $X_{i2} \in R^q$ denote two matrices, each set contains the same n samples from two different datasets. The maximum correlation coefficient between X_1 and X_2 is given by

$$\rho = max \frac{w_{X_1}^T X_1 X_2^T w_{X_2}}{\sqrt{\left(w_{X_1}^T X_1 X_1^T w_{X_1}\right)\left(w_{X_2}^T X_2 X_2^T w_{X_2}\right)}} \quad (13)$$

, where $w_{X_1} \in R^m$ and $w_{X_2} \in R^m$ are the canonical vectors and $m \le min(p,q)$ is the number of dimensions. The solution of CCA can be obtained by computing a generalized eigenvalue decomposition problem.

To improve CCA, DCCA is obtained from a subspace of input data sets which preserves both the data information and the class information. The given pairwise samples $D = \{(X_{i1}, X_{i2})^n_{i=1}\}$ are normalized and come from c classes. The standard CCA optimization is modified so that $X_1 X_2^T$ in Eq. (13) is replaced with $C_w - \eta C_b$ as follows;

$$\rho = max \frac{w_{X_1}^T \left(C_w - \eta C_b\right) w_{X_2}}{\sqrt{\left(w_{X_1}^T X_1 X_1^T w_{X_1}\right)\left(w_{X_2}^T X_2 X_2^T w_{X_2}\right)}} \quad (14)$$

, where C_b represents local between-class correlation and C_w represents local within-class correlation. η is an adjustable parameter that indicates the relative significance of C_b and C_w. The C_w is defined as;

$$C_w = \sum_{i=1}^{c} \left(X_k P_{n_i}\right)\left(X_l P_{n_i}\right)^T = X_k A X_l^T \quad (15)$$

$$X_k = \left[x_{k1}^{(1)}, \ldots, x_{kn_1}^{(1)}, \ldots, x_{k1}^{(c)}, \ldots, x_{kn_c}^{(c)}\right] \quad (15a)$$

$$X_l = \left[x_{l1}^{(1)}, \ldots, x_{ln_1}^{(1)}, \ldots, x_{l1}^{(c)}, \ldots, x_{ln_c}^{(c)}\right] \quad (15b)$$

$$P_{n_i} = \left[0, \ldots, 0; 1, \ldots, 1; 0, \ldots, 0\right]^T \in R^n \quad (15c)$$

, where n_i is the number of samples in the ith class of X_k or X_l, $x^{(i)}_{kj}$ is the jth sample in the ith class of X_k and $x^{(i)}_{lj}$ is the jth sample in the ith class of X_l. The P_{ni} is the new vector that contains one and zero elements.

One element represents the positions associated with class i. Therefore, the matrix A is defined as

$$A = \begin{bmatrix} 1_{n_1 \times n_1} & & & & \\ & \ddots & & & \\ & & 1_{n_i \times n_i} & & \\ & & & \ddots & \\ & & & & 1_{n_c \times n_c} \end{bmatrix} \in R^{n \times n} \quad (16)$$

, which contains class information. On the other hand, the C_b is defined as;

$$C_b = \left(X_k 1_n\right)\left(X_l 1_n\right)^T - X_k A X_l^T \quad (17)$$

Specifically, since X_k and X_l are mean-normalized, the $(X_k 1_n)(X_l 1_n)^T$ is zero. In [20,26], η can be omitted, the discriminate correlation matrix can be defined by:

$$C_w - \eta C_b = \left(1 + \eta\right) X_k A X_l^T = X_k A X_l^T \quad (18)$$

The solution when there are two datasets can be obtained by solving a generalized eigenvalue problem:

$$\begin{bmatrix} 0 & X_1 A X_2^T \\ X_2 A X_1^T & 0 \end{bmatrix} \begin{bmatrix} w_{X_1} \\ w_{X_2} \end{bmatrix} = \lambda \begin{bmatrix} X_1 X_1^T & 0 \\ 0 & X_2 X_2^T \end{bmatrix} \begin{bmatrix} w_{X_1} \\ w_{X_2} \end{bmatrix} \quad (19)$$

, where λ is the eigenvalue corresponding to the eigenvector (w_{X_1} and w_{X_2}). For multi-datasets $M > 2$, the projection canonical vectors and variables of multi-datasets are denoted as w_{Xm} and $z_m = w^T_{Xm} X_m$, $m = 1, \ldots, M$, where M is the number of sets and X_m is multi-datasets obtained from different sources. The feature combination of multi-feature sets (Z_M), which is the input to train in the classifier can be performed by:

$$Z_M = \frac{1}{M-1} \sum_{m=1}^{M} w_{X_m}^T X_m \quad (20)$$

3. The proposed algorithm

In this section, the proposed algorithm in sleep-wake detection and multi-class sleep stage classification are described. This paper proposes two main points. The first point, we proposed the hybrid dimensionality reduction to improve the automatic sleep stage classification by combining the feature selection based on the NIAs and the feature transformation technique by DCCA/CCA as shown

in Fig1. The second point, we proposed the novel BGWO called the dissimilarity binary grey wolf optimization - canonical correlation analysis (DisBGWO-CCA), and the dissimilarity binary grey wolf optimization - discriminate canonical correlation analysis (DisBGWO-DCCA) as explained in Fig. 2. The procedure is as following: the first step, all datasets are divided into training, validating and testing sets and the feature position of each dataset is generated by random binary values {0,1}. The value 1 means the feature is selected and 0 means not selected. The second step, the selected features of each set are fused by the CCA or DCCA techniques. The different sets are pairwise matched as (X_1, X_2) by the combination technique. The total member of pairwise is $K(K-1)/(2\times1)$, where K is the total number of feature sets. The DisBGWO-CCA aims to find the pair of projections by Eq.(13), while the case of the projection sets of DisBGWO-DCCA is calculated by Eq.(14) and the eigenvectors are solved by Eq.(19). In fusion procedure, the feature fusion of the multi-feature sets can be performed by Eq.(20). After that, the fused feature vector is evaluated by the fitness function, which consists of the maximum classification accuracy and the minimum features. The fitness function (F) is represented as:

$$F = \gamma\left(\frac{t}{T}\right)\times100 + (1-\gamma)\left(\frac{f_\alpha - f_s}{f_\alpha}\right) \qquad (21)$$

, where t is the number of correctly classified instances, T is the total number of instances, f_s is the number of fused features, f_a is the number of the total features and γ is the weight constant to control the importance of classification accuracy and the number of features. The third step, the feature positions, $x_d(t+1)$, are updated by BGWO technique using Eq.(10), and then the optimal feature sets are fused and evaluated using CCA/DCCA and the fitness function, respectively. After that, the fitness functions are ranked and the three best solutions are selected. The 1st best solution is the binary vector, $S_1 = \{x_{11},x_{12},\ldots,x_{1d}\}$, where d^{th} dimension bit. The 2nd best and the 3rd best solutions are $S_2 = \{x_{21},x_{22},\ldots,x_{2d}\}$ and $S_3 = \{x_{31},x_{32},\ldots,x_{3d}\}$, respectively. Next step, a mutant solution (S_m) is create by three best solutions, S_1, S_2 and S_3. The S_m can be defined as follow:

$$S_m = S_1 + \phi(S_2 - S_3) \qquad (22)$$

, where ϕ is the scaling factor. To transform the Eq. (22) into the concept of Jaccard coefficient

dissimilarity in section 2.2. Eq.(22) is rewritten in the form of Eq.(23).

$$S_m - S_1 = \phi(S_2 - S_3) \qquad (23)$$

Then, the "-" operator of Eq.(23) is substituted with a dissimilarity (*Dis*) measuring the difference between binary vectors, so it is reformulated as;

$$Dis(S_m, S_1) \approx \phi \times Dis(S_2, S_3) \qquad (24)$$

We use the symbol "$\approx$" instead of "$=$", because the dissimilarity between S_m and S_1 should be close to the result of $\phi\times Dis(S_2,S_3)$ as much as possible. In the S_m solving, the problem optimization can be represented as;

$$min\{Dis(S_m, S_1) - \phi Dis(S_2, S_3)\} \qquad (25)$$

Input: Labelled datasets, Number of dimensions, Maximum of iterations(*MaxInt*), The parameters of BGWO ($\alpha = 10$, $c = 0.5$ and $\phi = 0.8$), The weight constant of fitness function ($\gamma = 0.8$).
Output: The best fitness, the best solution.
Step 1: Divide all datasets into training, validation and testing sets and initialize population randomly $n_i = (1,...,N)$.
   : Generate the feature positions of each dataset x_i^j ($t=0$) where $j = (1,...,d)$.
Step 2: Fuse data sets using CCA/DCCA technique and evaluate using the classification accuracy and compute the fitness $F(i)$ by Eq.(21).
Step 3: While t < *MaxInt*
   : Update new solutions of each dataset (*new wolves*) by BGWO in Eq.(10).
   : Fuse data sets using CCA/DCCA technique by Eq.(13,14).
   : Evaluate using the fitness F(*new wolves*) by Eq.(21).
   : Compare $F(i)$ and F(*new wolves*)
   : If F(*new wolves*)$>F(i)$, replace $F(i)=F$(*new wolves*) *and* $n(i)=$ *new wolves*(i)
   : Find the three best solutions.
Step 4: Generate mutation solution of each feature set, S_m by Eq. (22) using dissimilarity Jaccard coefficient.
   : Fuse feature sets using CCA/DCCA and evaluate the fitness $F(m)$ by Eq.(21).
Step 5: Compare $F(i)$ and $F(m)$
   : If $F(m)> F(i)$,
     replace $F(i)= F(m)$ *and* $n(i)=Mutation$ (i)
Step 6: Rank $F(i)$ and update the best global solution.
Step 7: Obtain the best solution and run the process until the stop criteria are reached.

Figure. 2 The DisBGWO-CCA and DisBGWO-DCCA algorithms for multi-sets

The S_m is determined by using mathematical programing as found describing in [13]. After that, the mutant solution is fused using CCA/DCCA and the fitness function in Eq.(21) is used to evaluate. The fitness function of X_m solution (F_m) is compared with the current best fitness function of F_i, if the F_m better than F_i, the position of mutation solution will replace that. In each iteration, the best position and the best fitness are expected to be found. Finally, the process is stopped when the number of iterations reaches the maximum defined by the user.

4. Experiment design

4.1 Datasets

This experiment uses the data sets of sleep analysis or PSGs, which were recorded at the Sleep Laboratory of André Vésale Hospital (Belgium) [27]. These contained recorded signals and sleep staging labeled by sleep experts. The PSG datasets consisted of healthy adult subjects and patients with various pathologies. This database was acquired by using a standard digital 32 channels. We employed three different signals, namely EEG signals (CZ-A1, FP1-A1, and O1-A1), EOG signals (right outer canthus ROC and left outer canthus LOC), and single-lead ECG signal. Then these signals were segmented with a 30 seconds lengths based on the AASM criteria. Our study obtained 49,978 datasets in total with the following sleep stage labels: 10,862 datasets from the waking stage, 9,521 datasets from the REM stage, 8,941 datasets from the NREM1 stage, 9,557 datasets from the NREM2 stage and 11,097 datasets from the NREM3 stage.

4.2 Features extraction

The EEG, EOG, and ECG signals are extracted because these signals provide important information about the neuronal activity of the body during sleep. In our experiment, the multiple datasets are represented by the extracted features of six channels of the PSG recording, containing three EEG channels, two EOG channels, and a single ECG channel. The number of original features is 461 features (243 EEG features, 162 EOG features, and 56 ECG features). The following subsections describe the feature extraction of these signals.

4.2.1. EEG and EOG feature extractions

Each EEG placement in this study corresponds to the basic of the cerebral cortex and is the sub-position of the standard 10-20 system introduced by the AASM criteria. The various frequency bands of the EEG signal such as delta, alpha, theta, and beta represent the brain function. In this study, the EEG features are considered in domains of time, frequency, and time-frequency. The time domain is calculated by simple statistical techniques, while the power spectrum density of EEG frequency bands is estimated by the Fast Fourier Transform (FFT) and the Welch techniques. Moreover, the discrete wavelet transform (DWT) method is used to provide different frequency bands with coarse approximations and detailed information in the time-frequency domain.

In addition, the EOG signal including the ROC and LOC electrode positions is suggested to record the eye movements. This signal is used to define sleep stage individually based on the AASM standard. The processing of this extracted feature is similar to the EEG procedure. Details of the procedure of the suitability of all features can be found in [6] and other recent studies [28, 29].

4.2.2. ECG feature extraction

Several literatures [30-31] have suggested that the ECG is a standard solution to estimate the sleep stages. These studies utilized the heart rate variability analysis (HRV) that was derived from R-peak to R-peak intervals of the ECG signal to evaluate the balance of the autonomic nervous system (ANS) during sleep. The parameters of HRV consist of four domains such as time domain, frequency domain, time-frequency domain and nonlinear domain. In spectral domain processing, the features are calculated by power spectrum density (PSD) using the spectral estimation methods including fast fourier transform (FFT) and autoregression (AR) techniques. In addition, some nonlinear techniques such as correlation analysis, multi-scale entropies, and detrended fluctuation analysis (DFA) are employed to show significant differences of sleep stages [32]. The overall feature extractions are presented in [6].

4.3 Parameter setting

To show the performances of the proposed algorithm, the following algorithms are demonstrated: BPSO-CCA/DCCA, BCS-CCA/DCCA, EBCS-CCA/DCCA, BGWO-CCA/DCCA and DisBGWO-CCA/DCCA. The proposed algorithms are also compared with other conventional algorithms, which use only technique of feature selection or feature extraction such as SFS, CCA, DCCA, and BGWO. In the experiment, the parameters as shown in Table 1 are used.

284

Table 1. The parameters setting

Algorithms	Parameters
BPSO - CCA/DCCA	Acceleration factor related to *gbest* ($c_1 = 1$), *lbest* ($c_2 = 2$), and the weight constant ($w = 0.72$).
BCS - CCA/DCCA	The levy distribution (λ) = 1.25, the step size scaling weigh (α) = 1, and the probability of abandon (P_a) = 0.25.
EBCS - CCA/DCCA	The number of mutation solution ($N_{mut} = 10$), $\lambda = 1.25$ and $\alpha = 1$.
BGWO - CCA/DCCA	The parameters of the sigmoid function ($m = 10$, $n = 0.5$).
DisBGWO - CCA/DCCA	$m = 10$, $n = 0.5$, and the scaling factor (ϕ) = 0.8.

The parameters of BPSO are selected as in [9] such as $C_1 = 1$, $C_1 = 2$, and $w = 0.72$. The value of step size scaling factor (α) of original BCS and EBCS is set to 1 that is chosen to be large enough to generate new solutions as in [10, 24]. Levy distribution (λ) and the probability (P_a) of worst nest abandonment are also set to 1.25 and 0.25, respectively. The sigmoid parameters (m and n) of BGWO are set to 10 and 0.5 as in [12]. The scaling factor (ϕ) of DisBGWO is experimentally tuned using trial and error with the result of 0.8. The $\gamma = 0.8$ is used in the fitness function that the classification performance is more important than the number of features. The population size and the number of maximum iteration for all algorithms are set to 30 and 300, respectively. In feature selection processing, three different methods including small, random, and large initializations are employed to initialize a solution of feature selection. Detail of these methods can be reviewed in [14]. Furthermore, the fitness function is considered by k – fold cross validation to verify the reliability. We used the K-nearest neighbors ($k = 5$) to evaluate the algorithm performance because this classifier has the ability in terms of less time computation and higher classification accuracy than other algorithms namely random forest, decision tree, and support vector machine relied on our previous experiments. After the training phase, the best solution of each algorithm is used on the test phase to get the performance. The following measures are used to evaluate the algorithm performance:

1. The average of accuracy (*AvgAcc.*) is the ratio of the total number of the correct prediction and the total number of the labels. It is determined using

$$ArgAcc = \frac{1}{K}\sum_{k=1}^{K}\left(\frac{TP+TN}{TP+FP+FN+TN}\times100\%\right)_k \quad (27)$$

, where K is the number of times for testing, TP is the number of correct predictions in the positive instances, FP is the number of incorrect predictions in the positive instances, FN is the number of incorrect predictions in the negative instances, and TN is the number of correct predictions in the negative instances.

2. The average of sensitivity (*AvgSent.*) is the average of the ratio of positive instances that were correctly classified, calculated as

$$AvgSent = \frac{1}{K}\sum_{k=1}^{K}\left(\frac{TP}{TP+FN}\times100\%\right)_k \quad (28)$$

3. The average of specificity (*AvgSpec.*) is the average of the ratio of the negative that was correctly classified, calculated by

$$AvgSpec = \frac{1}{K}\sum_{k=1}^{K}\left(\frac{TN}{TN+FP}\times100\%\right)_k \quad (29)$$

4. Standard deviation (*Std.*) represents the variance of running to give the best solutions for T times. The *Std* can be defined as

$$Std = \sqrt{\frac{1}{K}\sum_{k=1}^{K}\left(p_k - mean(p)\right)} \quad (30)$$

The p_k is the best solution obtained from the kth run. All experiments were carried out using the MATLAB program on a personal computer (Intel Core i5-2430M CPU, 2.4 GHz, with 8 Gb RAM. running Windows 8.1 as the operating system).

5. Experiment results and discussion

This experiment evaluates the effectiveness of the proposed algorithm for the wake-sleep detection and the multi-class sleep stage classification. All datasets are randomly divided into three groups including training, validation and testing datasets.

5.1 Results on the training phase

Fig.3 (a-b) shows the performance of the combination between NIAs and CCA/DCCA using small, normal and large initializations to find the best method to initialize a solution of the feature selection on sleep-wake detection and multi-class sleep stage classification experiments. The performance of the DisGWO-DCCA achieves the best in case of large initialization, where all positions are set to 1 after that the optimization will search features to remove for increasing the fitness value. Therefore, the large initialization can globally search the optimal features better than the other initialization methods.
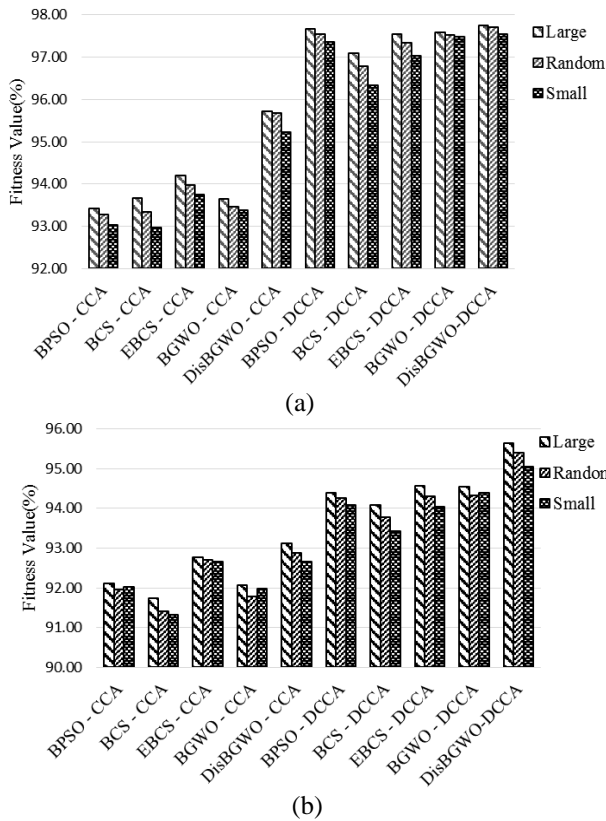
(a)

(b)

Figure. 3 The fitness value of the different algorithms using small, random and large initializations on: (a) sleep-wake detection and (b) multi-class sleep stage
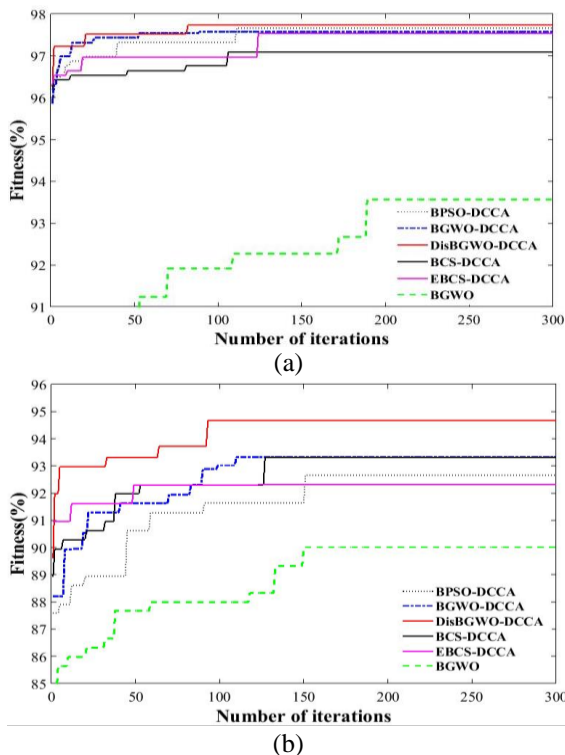

(a)

(b)

Figure. 4 Fitness comparison algorithms on: (a) sleep-wake and (b) multi-class sleep stage classification

We used *k*-fold cross-validations, including *k* = 2, 4, and 5, to manage the training dataset and validation set appropriately. Table 2 and 3 show the results of fitness values of 14 different approaches, which are the conventional feature selection methods and the proposed algorithms, on sleep-wake detection and multi-class sleep stage classification. The result of sleep-wake detection shows that the best performance is obtained from the DisBGWO -

Table 2. Comparative fitness values on training sets of sleep-wake detection experiment.

Algorithms	Fitness value (%±*std*)		
	2-foldCV	4-foldCV	5-foldCV
Existing algorithm			
SFS	88.94±0.71	89.97±0.72	91.34±0.36
CCA	88.04±1.04	88.72±0.51	90.98±0.54
DCCA	90.05±0.57	91.01±0.76	91.47±0.68
BGWO	91.98±0.62	92.14±0.47	93.02±0.68
Proposed algorithm			
BPSO-CCA	92.98±0.34	93.21±0.37	93.42±0.84
BCS-CCA	93.19±0.51	93.21±0.51	93.67±0.54
EBCS-CCA	93.21±0.62	93.26±0.52	94.21±0.53
BGWO-CCA	93.37±0.38	93.53±0.44	93.65±0.62
DisBGWO-CCA	**94.03±0.49**	**94.55±0.38**	**95.72±0.45**
BPSO-DCCA	96.68±0.32	96.93±0.38	97.59±0.33
BCS-DCCA	96.52 0.39	96.54±0.60	97.09±0.46
EBCS-DCCA	96.53±0.51	96.67±0.69	97.54±0.52
BGWO-DCCA	96.70±0.31	96.98±0.29	97.58±0.49
DisBGWO-DCCA	**96.74±0.16**	**97.35±0.16**	**97.75±0.32**

Table 3. Comparative fitness values on training sets of multi-class sleep classification.

Algorithms	Fitness value (%±*std*)		
	2-foldCV	4-foldCV	5-foldCV
Existing algorithm			
SFS	81.97±0.16	82.78±1.21	84.47±0.43
CCA	83.97±0.64	85.47±0.34	86.72±0.74
DCCA	84.34±0.77	87.00±0.97	87.44±0.67
BGWO	87.42±0.27	88.45±0.37	89.74±0.51
Proposed algorithm			
BPSO-CCA	91.67±0.59	92.18±0.64	92.10±0.95
BCS-CCA	90.35±0.52	90.53±0.41	91.73±0.44
EBCS-CCA	90.87±0.73	90.88±0.73	91.92±1.01
BGWO-CCA	89.58±1.42	92.06±0.69	92.07±0.67
DisBGWO-CCA	**91.98±0.48**	**92.85±0.37**	**92.76±0.24**
BPSO-DCCA	93.06±0.82	93.15±0.98	94.38±0.95
BCS-DCCA	93.41±0.65	93.35±0.13	94.09±0.44
EBCS-DCCA	93.97±0.47	94.11±0.56	94.55±0.98
BGWO-DCCA	92.82±0.76	93.26±0.49	94.54±0.58
DisBGWO-DCCA	**94.92±0.32**	**95.25±0.29**	**95.85±0.34**

Table 4. Comparative accuracy values on testing sets of sleep-wake detection

Algorithms	Sent.(%)	Spec. (%)	Acc. (%)
Existing algorithm			
SFS	79.57±1.33	87.47±1.13	83.06±0.66
CCA	81.12±1.05	89.27±0.64	85.97±0.57
DCCA	82.72±0.57	90.45±0.63	86.34±1.01
BGWO	83.03±0.86	92.41±0.71	88.04±0.81
Proposed algorithm			
BPSO-CCA	83.52±0.14	95.79±0.75	89.66±0.56
BCS-CCA	84.78±1.53	92.24±0.72	88.51±0.94
EBCS-CCA	87.83±1.15	92.82±0.61	90.32±0.85
BGWO-CCA	86.30±1.72	92.58±0.24	89.44±0.89
DisBGWO-CCA	**89.19±1.03**	92.07±0.74	**90.63±0.48**
BPSO-DCCA	88.80±1.41	97.99±0.54	93.40±1.04
BCS-DCCA	87.95±1.25	97.77±0.91	92.86±1.10
EBCS-DCCA	88.40±1.06	97.88±0.57	93.14±1.68
BGWO-DCCA	88.62±1.61	97.94±0.49	93.28±0.77
DisBGWO-DCCA	**89.74±0.83**	**98.29±0.57**	**94.02±0.61**

Table 5. Comparative accuracy values on testing sets of multi-class sleep stage classification

Algorithms	Sent.(%)	Spec. (%)	Acc. (%)
Existing algorithm			
SFS	85.27±0.26	92.76±1.03	90.64±0.72
CCA	85.07±1.21	93.45±0.97	90.68±0.37
DCCA	86.42±0.87	93.07±0.71	91.04±0.45
BGWO	88.97±0.75	94.46±0.44	92.67±0.59
Proposed algorithm			
BPSO-CCA	88.57±1.18	92.77±1.15	90.67±1.35
BCS-CCA	86.94±1.42	93.80±0.46	90.38±0.73
EBCS-CCA	89.28±0.68	92.37±0.37	90.83±0.42
BGWO-CCA	90.20±1.74	92.75±0.81	91.47±0.92
DisBGWO-CCA	**90.66±0.77**	**93.08±0.77**	**91.87±0.37**
BPSO-DCCA	91.47±1.19	97.02±0.62	94.23±1.01
BCS-DCCA	92.01±1.59	97.46±0.87	94.74±1.17
EBCS-DCCA	92.32±1.28	97.08±0.69	94.70±0.83
BGWO-DCCA	91.83±1.33	98.08±0.63	94.95±0.35
DisBGWO-DCCA	**93.70±0.94**	**96.79±0.51**	**95.89±0.45**

DCCA algorithm in *k*-fold cross validations. The average fitness values are; 2-fold CV= 96.74 ± 0.16 %, 4-fold CV= 97.35 ± 0.16 %, and 5-fold CV=97.75 ± 0.32%. In addition, the result of multi-class sleep stage classification shows that the proposed algorithm provides the best performance with average fitness values as following; 2-fold CV = 94.92 ± 0.32 %, 4-fold CV = 95.25 ± 0.29 % and 5-fold CV = 95.85 ± 0.34 %. The best accuracy on various *k* cross validations is obtained from the proposed algorithm, which can perform better for classification tasks. Fig. 4 (a-b) shows the fitness

function curves of proposed algorithms and the best existing algorithm which is the BGWO as shown in Table 2 and 3. The figure shows that the DisBGWO - DCCA algorithm uses the least number of iterations that means converging to the best solution faster than other algorithms. In comparing with BGWO, the proposed algorithm can search the optimal features in both global and local space better than BGWO because DisBGWO - DCCA has created the mutant solution by three best solutions instead of the random search solution in space. Moreover, the proposed algorithm can efficiently combine different features.

5.2 Results on the training phaseResults on the testing phase

Since the best model has been constructed, it can be applied in testing phase to predict the class of previously unseen data. This experiment aims to measure the performance of the models. We evaluate the models based on different evaluation criteria including accuracy, sensitivity, and specificity. Table 4 and 5 present the comparative performance of ten different algorithms in sleep wake detection and multi-sleep stage classification. We can see from the tables that the DisBWGO-DCCA also performs with the highest performance (accuracy = 95.89 ± 0.45 %, sensitivity = 93.60 ± 0.94%, and specificity = 96.79 ± 0.51% in sleep-wake detection, and accuracy = 94.02 ± 0.61 %, sensitivity = 89.74 ± 0.83%, and specificity = 98.29 ± 0.57% in multi-sleep stage classification). In the testing process, this results can be concluded that the proposed model with DCCA has ability to classify sleep stage more efficiently because the class information is taken into consideration. So, the DCCA has an advantage in solving the weak relation between selected features obtained from different sources.

5.3 Evaluation of the proposed algorithm by the global data

To evaluate the proposed algorithm on the global data, the experiment used the test datasets of healthy subjects and sleep disorder patients to demonstrate the performance of the DisBGWO-DCCA approach in wake-sleep detection and multi-class sleep stage classification. Fig.5 (a-l) shows the box plots of average accuracies, sensitivities, and specificities of each group in the both experiments. The results also give high accuracy when testing with groups of healthy subject and patients.

From the experiment results, we can conclude that the proposed algorithm not only provides better
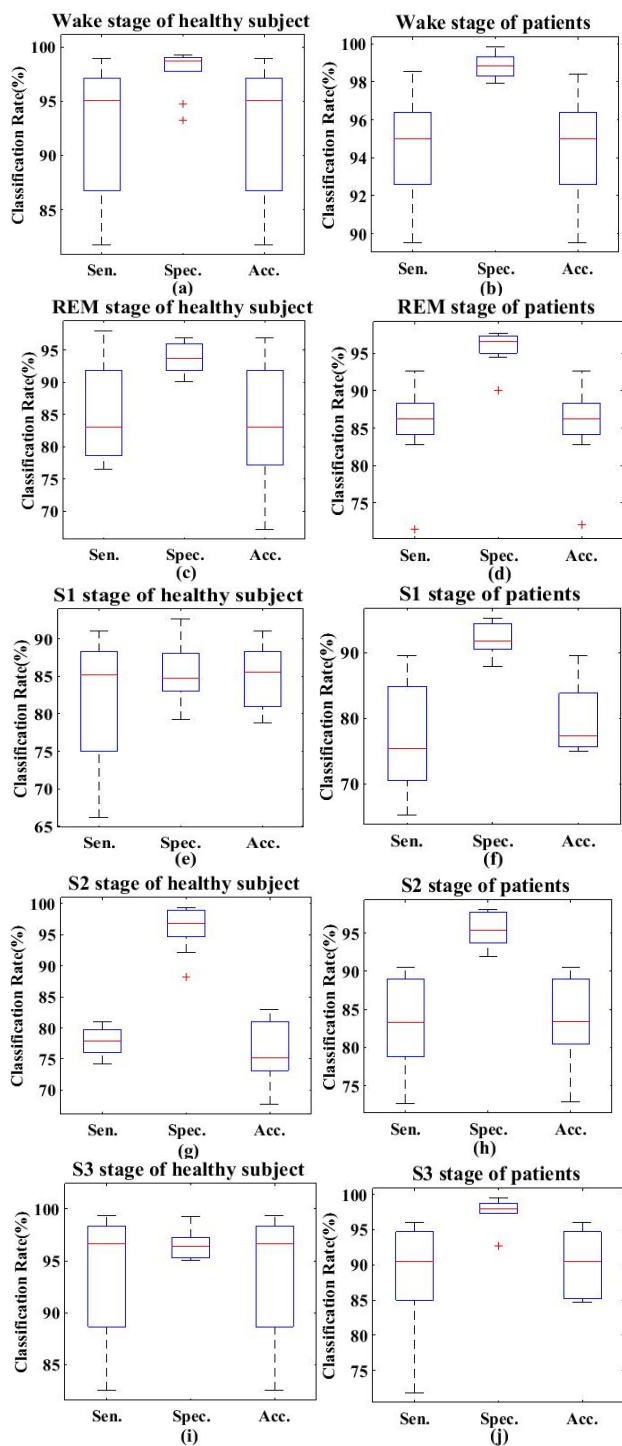
Figure. 5 Boxplot of the performance of healthy subjects and patients in: (a-b) wake stage, (c-d) REM stage, (e-f) S1 stage, (g-h) S2 stage, and (i-j) S3 stage

solution quality but also reduces computational cost, which also helps solving the computational time of the wrapper feature selection problem in training process. In addition, the proposed algorithm achieves the best classification performance in both sleep-wake detection and multi-class sleep stage classification when comparing with the conventional techniques.

6. Conclusion

This paper proposes and evaluates the hybrid dimensionality reduction for improving the automatic sleep stage classification in terms of high accuracy and less computation time. Our proposed technique is the first original approach that combines the NIAs and CCA/DCCA methods to reduce dimension from multi-dataset. The proposed framework consists of two main steps including the feature selection and the multi-sets transformation. Firstly, a binary NIAs, which is well known on feature selection methods such as BPSO, BCS, EBCS, BGWO and DisBGWO, are used to optimize features in each measurement of signals. Secondly, the effective CCA/DCCA performs feature combining to be inputs of the machine learning. In this paper, multi EEG signals, multi EOG signals, and single ECG signal are represented as multi-sets for evaluating and comparing the performance of hybrid dimensionality reduction algorithms. The evaluation process is performed using the criteria of evaluation to estimate the different aspects of the proposed system. From the experiment, the DisBGWO-DCCA, which is modified by using the differential evolution based on the dissimilarity of the Jaccard coefficient, outperforms the other methods. Moreover, the proposed algorithm can clearly classify the states of sleep on both sleep-wake detection and multi-class sleep stage classification when comparing with the previous studies [1-6]. In the future work, we plan to apply the proposed algorithm in the automatic sleep stage classification system and we will develop this algorithm in other applications of multi-sets.

References

[1] M. O. Mendez, M. Matteucci, and V. Castronovo, "Sleep staging from Heart Rate Variability: time-varying spectral features and Hidden Markov Models", *International Journal of Biomedical Engineering and Technology*, Vol. 3, No.3, pp. 246-263, 2010.

[2] A. Martin, G. Guerrero-Mora, G. Dorantes-Méndez, A. Alba, M.O. Méndez, and I. Chouvarda, "Non-linear analysis of EEG and HRV signal during sleep", In: *Proc. of International Conf. on Biomedical and Health Information*, pp.4174-4177, 2015.

[3] J.R. Yeh, C.K. Peng, M.T., S.C. Chen, C.Y. Wang, P.L. Lee, and J.H. Kang, "Investigating the interaction between heart rate variability and sleep EEG using nonlinear algorithms", *Journal of Neuroscience Methods*, Vol. 219, No.2, pp. 233-239, 2013.

[4] L. Zoubek, S. Charbonnier, S. Lesecq, A. Buguet, and F. Chapotot, "Feature selection for sleep/wake stages classification using data driven methods", *Journal of Biomedical Signal Processing and Control*, Vol.2, pp.171-179, 2007.

[5] S. Khalighi, T. Sousa, and U. Nunes, "Automatic sleep staging: A computer-assisted approach for optimal combination of features and PSG channels", *International Journal of Expert System with Application*, Vol.40, No.17, pp.7046-7059, 2013.

[6] P. Moeynoi and Y. Kitjaidure, "Dimension reduction based on Canonical Correlation Analysis technique to classify sleep stages of sleep apnea disorder using EEG and ECG signals", In: *Proc. of International Conf. on Electrical Engineering/ Electronic, Computer, Telecommunication and Information Technology*, pp.455-458, 2017.

[7] E. Pamukcu, H. Bozdogan, and S. Calik, "A novel Hybrid Dimension Reduction Technique for Undersized Information Complexity Criterion for Cancer Classification", *Journal of Computational and Mathematic Methods in Medicine*, Vol. 2015, pp. 1-14, 2015.

[8] R. Aziz, C.K. Verma, and N. Srivastava, "A novel approach for dimension reduction of microarray", *Journal of Computational Biology and Chemistry*, Vol.71, pp.161-169, 2017.

[9] B. Xue, S. Nguyen, and M. Zhang, "A New Binary Particle Swarm Optimization Algorithm for Feature Selection", In: *Proc. of International Conf. On the Applications of Evolutionary Computation*, pp.501-513, 2014.

[10] D. Rodrigues, L.A.M. Pereira, T.N.S. Almeida, J.P. Papa, A.N. Souza, C.C.O. Ramos, and X.S. Yang, "BCS: A Binary Cuckoo Search Algorithm for Features Selection", In: *Proc. of IEEE international Symposium on Circuits and Systems*, pp.465-468, 2013.

[11] E. Hancer, B. Xue, D. Karaboga, and M. Zhang, "A binary ABC algorithm based on advanced similarity scheme for feature selection", *Journal of Applied Soft Computing*, Vol. 36, pp.334–348, 2015.

[12] E. Emary, H.M. Zawbaa, and A.E. Hassanien, "Binary grey wolf optimization approaches for feature selection", *Journal of Neurocomputing*, Vol. 172, pp.371–381, 2016.

[13] Q. Li, H. Chen, H. Huang, X. Zhao, Z.N. Cai, C. Tong, W. Liu, and X. Tian, "An Enhanced Grey Wolf Optimization Based Feature Selection Wrapped Kernel Extreme Learning Machine for Medical Diagnosis", *Journal of Computational and Mathematical Methods in Medicine*, Vol.2017, pp. 1-15, 2017.

[14] U.G. Mahgri, S. Samanta, S. Das, and P.R. Chawdhury, "A Survey of Decision Fusion and Feature Fusion Strategies for Pattern Classification", *Journal of Institution of Electronics and Telecommunication Engineers Technical Review*, Vol.27, No.4, pp.293-307, 2010.

[15] W. Zuobin, M. Kezhi, and G.W. Ng, "Effective feature fusion for pattern classification based on intra-class and extra-class discriminative correlation analysis", In: *Proc. of International Conf. on the Information Fusion*, 2017.

[16] F. A. Faria, J. A. D. Dantons, A. Rocha, and R.D. Torres, "A framework for selection and fusion of pattern classifiers in multimedia recognition", *Journal of Pattern Recognition*, Vol.39, pp.52-64, 2014.

[17] T. Sun, S. Chen, J. Yang, and P. Shi, "A new method of combined feature extraction for recognition", In: *Proc. of International Conf. on Data mining*, pp.15-19, 2008.

[18] N.M. Correa, T. Eichele, T. Adali, Y.O. Li, and V.D. Calhoun, "Multi-set canonical correlation analysis for the fusion of concurrent single trial ERP and functional MRI", *Journal of Neuro. Image*, Vol.50, No.4, pp.1438-1445, 2010.

[19] D.R. Hardoon, S. Szedmak, and J.S. Taylor, "Canonical Correlation Analysis: An Overview with Application to Learning Methods", *Journal of Neural Computation*, Vol.16, No.12, pp.2639-2664, 2004.

[20] Q.S. Sun, S.G. Zeng, Y. Liu, P.A. Heng, and D.S. Xia, "A new method of feature fusion and its application in image recognition", *Journal of Pattern Recognition*, Vol.38, No.12, pp.2437-2448, 2005.

[21] Y.H. Yuan, Q.S. Zhou, and D.S. Xia, "A novel multi-sets integrated canonical correlation analysis framework and its application in feature fusion", *Journal of Pattern Recognition*, Vol.44, No. 5, pp.1031-1040, 2011.

[22] J. Via, I. Santamaria, and J. Perex, "A learning algorithm for adaptive canonical correlation analysis of several data sets", *Journal of Neural Networks*, Vol.20, No.1, pp.139-152, 2007.

[23] Y. Yuan, C. Ma, and D. Pu, "A novel discriminant minimum class locality preserving canonical correlation analysis and It applications", *Journal of Industrial and Management Optimization*, Vol.12, No.1, pp.251-268, 2016.

[24] S. Salesi and G. Cosma, "A Novel Extended Binary Cuckoo Search Algorithm for Feature

Selection", In: *Proc. of International Conf. on Knowledge Engineering and Applications*, pp.6-12, 2017.

[25] P. Jaccard, "The distribution of the flora in the alpine zone", *Journal of New Phytol*, Vol. 11, No.2, pp.37–50, 1912.

[26] X. Zhang, N. Guan, Z. Luo, and L. Lan, "Discriminative Locality Preserving Canonical Correlation Analysis", In: *Proc. of International Conf. on Pattern Recognition*, pp.341-349, 2012.

[27] S. Devuyst, T. Dutoit, J.F. Didier1, F. Meers, E. Stana, P. Stenuit, and M. Kerkhofs, "Automatic Sleep Spindle Detection in Patients with Sleep Disorders", In: *Proc. of International Conf. on IEEE Engineering in Medicine and Biology Society*, pp. 3883-3886, 2006.

[28] K.A.I. Aboalayon, M. Faezipour, W.S. Almuhammadi, and S. Moslehpour, "Sleep Stage Classification Using EEG Signal Analysis: A Comprehensive Survey and New Investigation", *Journal of Entropy*, Vol.18, No.9, pp.1-31, 2016.

[29] I. Zhovna and I.D. Shallom, "Automatic detection and classification of sleep stages by multichannel EEG signal modelling", In: *Proc. of International Conf. on IEEE Engineering on Medicine and Biology Society*, 2008.

[30] P.K. Stein and Y. Pu, "Heart rate variability, sleep and sleep disorders", *Journal of Sleep Medical Review*, Vol.16, No.1, pp.47-66, 2012.

[31] E. Tobaldini, L. Nobili, S. Strada1, K. R. Casali, A. Braghiroli, and N. Montano, "Heart rate variability in normal and pathological sleep", *Journal of Frontiers in Physiology*, Vol.16, 2013.

[32] V. Steven, "Heart Rate Variability: linear and non-linear analysis with applications in humans physiology", *Doctoral Thesis of Faculty of Electrical Engineering*, 2010.