International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 22 (2018) pp. 15908-15915

Inherent multi-view feature fusion using canonical correlation for classification of Electroencephalogram (EEG) signals

Arup Sarmah
Research Scholar
Department of Computer Science
Gauhati University
Guwahati, Kamrup, Assam, India

Dr. Sanjib Kalita
Assistant Professor
Department of Computer Science
Gauhati University
Guwahati, Kamrup, Assam, India

Dr. B K Dev Choudhury
Associate Professor & Principal
Pub Kamrup College
Baihata Chariali, Kamrup, Assam, India

ABSTRACT

Improvement of a feasible support system for automating staging of neural disorder based on single channel Electroencephalogram (EEG) is vital to speed-up diagnosis process by lightening the load of the clinician of analyzing large volume data and to accelerate large scale research. Most of the prior works arrange features elicited from precise distribution of signal or yield poor performance. In this article, inherent multi-view feature fusion using Canonical Correlation Analysis (CCA) for automatic staging of EEG signals is proposed. Segmenting EEG signal uniformly, we formulate multi-view or feature matrices (FMs) for all class distributions. Discrete wavelet transform (DWT) has Also been applied on FMs to extract their statistically independent wavelet FMs. Afterwards, CCA analysis is performed to extract low order features, which are then fused via parallel and serial fusion. Finally, two global descriptors are derived for classification. Finally Statistical test is performed to validate effectiveness of selected feature space. We then employed various classification models and investigated accordingly for various models. Experimental outcomes make clear that the performance of proposed feature fusion based automatic staging algorithm is superior to many state-of-the-art ones.

Keywords: Electroencephalogram (EEG), Feature Fusion, Multivew Vectors, Canonical Correlation

1. INTRODUCTION

In essence, newly in pattern recognition community [1, 2] multi-views learning (MVL) is gathering much more consideration. It can embed multiple views' information (i.e., features) of a given object. The term view symbolizes set of data or feature matrix. Usually in MVL same object is represented by various views, which form different feature spaces. Feature spaces may have same or different statistical nature. Nevertheless associations of features from various views through a precise etiquette produce discriminate measure to boost the characterization of the object. Thus, it avoids feature biasing and dimensionality issues, common weaknesses in learning situation. Therefore, it can well-felicitate the learning task and endorses feasible implementations. Availability of multiple measures and diverse nature of data distribution are inherent to many engineering applications [3]. Moreover feature data set may enclose many redundancies and extraneous information. So, removal of redundant matters from the feature dataset before subjecting to proper selection of feature space is an indispensable provision to obtain feasible support system [4]. Feature fusion, which is believed to be most effective in any decision model notably, helps to derive compact representation of large scale information through dimension reduction strategy [5]. There are various successful applications such as extraordinary application includes image retrieval, video annotation and document clustering [1] and web classification [6] etc. . Multifarious advantages of such learning models motivate to address a frame work for automated staging of Electroencephalogram (EEG) signals.

2. PROBLEM DESCRIPTIONS

Efficient modeling to derive feasible algorithm that can detect the state of disease at early stage is vital and hence the emergent in medical profession [7, 8]. Now a day to enhance large scale biomedical research and speed up medical practice substantial attention is paid off [9]. These approaches can overcome the limitations of traditional approaches. For example detection and classification of neurological disorders at early stage is essential. Cause of diseases changes the anatomy and physiology of motor units (MUs) and thereby, depreciates the quality of lives. Therefore, research is going on for such algorithm pursuit. Such algorithm implementation in portable devices increases the reliability of users for home-care with easy access, which is one of the most emergent research directions in medical domain in recent time. Traditional assessment requires expensive human resources, Rater's level of expertise and experience. Further it fails to provide quantitative information about severity and improvement. Additionally, analysis involves multiple assessments which make manual conclusions difficult and prone to errors. Sometimes quick diagnosis is not possible. In urgent clinical cases and also, precludes the large-scale population studies. For that reason, a huge number of support systems are being anticipated. In most of prior approaches, learning patterns are extracted from specific frames of signal or dominated contents of pre-defined class distribution. Although such assumption makes the model simple, it may fail to offer vigorous learning framework. Mainly two reasons are there. Firstly, underlying data distribution of signal is complex and not pre-known. As a results feature extraction from specific frames of signal may not be appropriate. Secondly, signals like EMG, EEG etc., are non-stationary and non-linear, also it is highly subjective and more it relies on a) nature of disease b) sub-class of disease, and c) clinical setting [2]. Therefore, feature requested from multiple views associated with given class distribution could provide more discriminate ability. We proposed multi-view feature fusion based on canonical correlation analysis (CCA), referred as mCCA. In works of Lin et. al, each template is regarded as specific view and inherent multiple views' features are trained with classifier. It is believed that implementation of our algorithm in health care profession not only alleviating the onus of clinician of large volume data but also expedite the diagnosis research. Thus it could enable prevention of most of real life problems. With such reliable integration, it could corroborate the users for early diagnosis and routinely check-up without the intervention of a physician. Thus, it is becoming a part and parcel of engineered furnishings.

3. RELATED WORKS

3.1 CCA based learning

Data driven fusion of multimodality data is an especially puzzling problem since brain imaging data types are essentially different in nature, making it difficult to analyze them together without making a number of assumptions, most often impracticable about the nature of the data. Two major concerns in learning models are Feature biasing and large-dimensionality. The technique feature fusion can reduce the entire issues. Such systems have potential values for real-world applications [12, 13]. Dimensionality enhances the learning parameters which increases model complexity [14], while feature biasing downsides the reliability of suggestions. Even sometime it provides trivial consequence. Feature fusion strategy in such cases plays a vital role. It projects the original feature space to a well-defined co-ordinate system from where apt dimensional features are extracted [4] and then, use them for discriminate representations of pre-defined sets. Thus, it abridges the dimension of feature space by reducing its redundancy and irrelevant information and thus boosting the system recital. Extracted features have much robustness in real-life applications, especially when the best feature sets are unknown [15]. Two most broadly used feature projection are principal components analysis (PCA) [16, 17] and linear discriminant analysis (LDA) [18, 19]. Unlike PCA, LDA includes the class information and find the projection that best separate the class. Recently, CCA [20] learning is popular and are successfully used in many applications. It extracts unique feature set based on the degree of proximity between two set of vectors extracted from same or different objects. In the recent past, various QA methods defining various feature extraction techniques from EEG signals have been reported and subsequence results encourage for superior algorithm pursuit in addition to the significant reinforcement towards real-time implementations. For instance, Hassanpour et al. [27] developed time-frequency based seizure detection technique. In [28], the author developed a versatile DWT based technique. Hassan et al. [30] adopt a linear programming boosting (LPBoost). First, use ensemble empirical mode decomposition to decompose EEG and spectral moments are used for classification. Additionally, Bootstrap aggregating in [31], tunable-Q factor wavelet transforms (TQWT) [32], Adaptive Boosting and decision trees [33] also reported. Orhan et al. [34] introduces a neural network using DWT coefficients for epilepsy diagnosis. Soomro et al. [35] adopt the combination of canonical correlation (CCA) analysis and neural network for epileptic seizures prediction. Kiymik et al. [36] use power spectral density based approach to study alert, drowsy and sleep.

3.2 EEG classification

Over the decades literature gives us various support models. Such methods either used EEG features or its component motor unit action potential (MUAP) features. By using feature extraction techniques such as coefficients of autoregressive model (ARM) [29], frequency features [30], and wavelet features etc.EEG features are extracted.

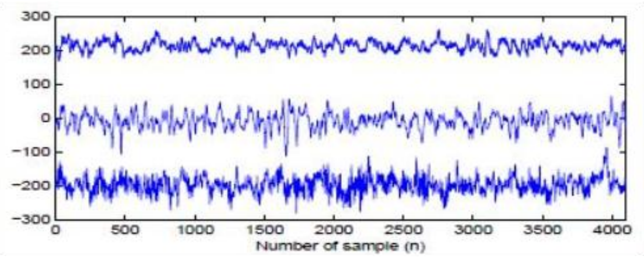


Fig 1: Three typical EEG pattern

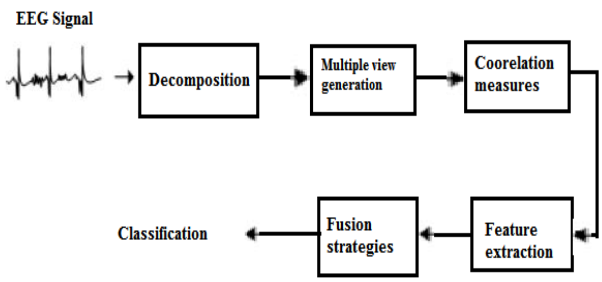


Fig 2: Block diagram

Freshly, CCA has emerged as a useful means of analyzing non-stationary modality signals such as EEG, EMG [23]. Dealing with large-scale information in term of compact views is the major advantage of CCA. Such compact view are essential for classification related problem and has been shown encouraging. However, there is some limitation in the literature on EEG classification using CCA. In the present article, an automatic staging of EEG signals using multi-view statistical feature fusion is proposed. To the best of our knowledge, use of CCA based fusion model for automated EEG is not reported in literature so far. This study therefore opens up new possibilities of research where multiple classifiers can be employed to further ameliorate the performance and to ensure diversity of measures. Diversity of measures ensures the integrity of this work and its possible extension for real-life applications. Further, the effectiveness of statistical features is demonstrated by computing p-values [46]. Finally, the performance of our approach is compared with those of the state-of-the-art methods. It is shown that in general our method achieve better performance and minimum multiplicity in measures.

4. MATERIALS

The algorithm is tested with publically available database of 150 recordings (50 A, 50 B and 50 E) [29]. Five sets of data are encoded from A to E, each of which contains 100 single channel EEG recordings and each collected for 23.6s duration. The data sets were recorded at the University Hospital Bonn, Germany with inbuilt amplifier and 12 ADC at sampling rate of 173.61Hz (i.e., n=fst=173.6 × 23.6= 4097 samples). Further, band setting of filter was 0–60 Hz. During data collection the healthy volunteers were relaxed in an awake state with eyes open (A) and eyes closed (B), respectively. Sets C, D, and E originated from our EEG archive of pre-surgical diagnosis. C D were measured during seizure free interval whereas set E only contained seizure activity.

5. METHOD

5.1 Feature extraction model

In this section, we develop mCCA based feature extraction model. Highly significant pairs of variables from CCA pair transformations are used as features and then, deployed two feature fusion techniques to derive discriminate patterns for classification of NDs. Fig.2 shows the schematic outline of adopted decision support framework. N-point EEG signal is segmented into a set of sequences $X=[x_1, x_2, x_3 \ldots x_p]$, where each sequence x_p have equal number of points. Next arranging d number of sequences in row-wise onto matrix template referred as view (i.e., feature matrix) we create a set of variables expressed mathematically as below:

$$Xi = [x_1; x_2; x_3; \ldots\ldots\ldots; x_{d-1}; x_d] \in R^{dxni} \qquad (1)$$

For given two such variables X and Y , CCA searches two sets of basis vectors, one for X and the other for Y , such that the correlations between the projections of the variables onto these basis vectors are mutually maximized. Consider linear transformation of components, also known as variates-

$$u = Ax_1x_1 + \cdots\cdots\cdots + Ax_kx_k = A_x^T X \qquad (2)$$

$$v = B_{y1}y_1 + \cdots\cdots + By_ky_k = B_y^T Y$$

CCA find weight vectors Ax = [Ax1;,,,,,;Axk] and By = [By1;,,,,,,,,,; Byk]that maximize the correlation ρ the variates u and v by solving following optimization problem [25]:

$$\max_{Ax,By} \rho(u,v) = \frac{E(uv)}{E(u^2)E(v^2)} = \frac{A_X^T C_{xy} B_y}{\sqrt{(A_X^T C_{xx} A_x)(B_y^T C_{yy} B_y)}} \qquad (3)$$

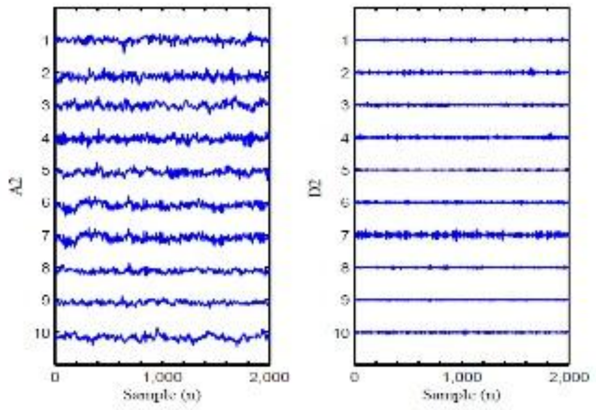
Where Cxx and Cyy the autocovariance matrices and Cxy the crosscovariance matrix of X and Y. However, to avoid large dimensionality and have computation efficient, we first perform PCA on the variables. This two-stage PCA+CCA approach reduce dimensionality as well as singularity issues [5]. Besides, mean of each row from the view matrices are removed to make centered data matrices. As defined in [19], this optimization problem (3) is solved by using an efficient Singular Value Decomposition (SVD) technique, which involves d×d dimensional matrix. Assume that X and X form unitary orthogonal bases for two linear subspaces. Let the SVD of $X^TY \in R^{dXd}$ be

$$X^TY = U\Lambda V^T = [U_1, U_2]\ \mathrm{diag}\ (\Lambda_d, 0)\ [V_1, V_2]^T = U_1\Lambda_d V_1 \qquad (4)$$

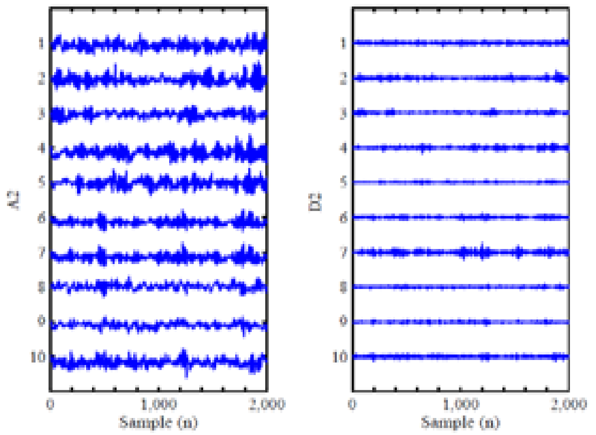
Where, U and V are two left and right singular orthogonal matrices of Cxy and Cyx, i.e., $UU^T = VV^T = I_d$. Further, mathematically it infers that U2 falls in the null space of X (i.e., $X^TU_2 = 0$), which indicates that it is uncorrelated component [24]. Additionally, tumbling the dimensionality based on the degree of similarity between two views, learning parameters can be further reduced, which in turn reduces the computational cost. Singular values represent canonical correlations, and the associated eigenvectors are given by

$$A = XU_1 = [A_1; \cdots; A_d];\ B = YV_1 = [B_1; \cdots; B_d] \qquad (5)$$

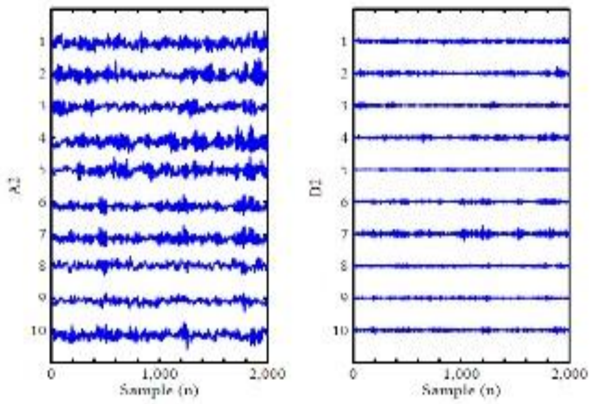
Diagonal elements of Λd that are in descending, measure the strength of feature vectors (FVs). Noteworthy factor is that the correlation between same indices pairs are none zero, due to orthogonal nature of FVs, e.g., correlation, r = 0, for pair A1;B2. Intuitively, first few pairs show significant proximate behaviour (Fig.4), i.e., they well-capture the idiosyncratic information from pair variable. Further, each FV of one set similar to corresponding FV in other set notwithstanding the data variation. Besides, in the proposed method over fitting of model and singularity of both scatter matrices are averted by dint of regularization [24]. Nonetheless, it does not mutate the projections as Eq.(3) independent of scaling to A and B. Low dimensional view in terms of A,B well preserves most of the energy contents by making use of subspace learning [19]. This strategy is applied to all pair of variables and subsequently estimated their projected vectors. The major challenge is a learning compact distribution of congenital information embedding to effectively explore the class information. Besides, to further avoid feature biasing DWT is performed on the variables and then, considering low frequency sub-band components frequency domain variables are generated for analysis. Fig.3 is second level wavelet decomposition. Choice of level and wavelet function depends on frequency content in the signal. Here, we use wavelet function Daubechies 2 wavelet (db2) with two vanishing moments [45]. Afterwards, using aforementioned method second statistically independent set of features is estimated. This multi-domain multi-view embedding could be the better probe than single frame or view embedding. It is worth indicating that the proposed method has relatively of high computational cost as compared to single view approach, which may presumably due to incorporation of multi-template selection strategy to extract relevant FVs for effective learning to have better diagnosis value. Thus, it is a trade-off between the qualities of feature space and computation cost. The use of CCA based features in our method has been motivated by its success in various applications such multi-modal analysis, image recognition, myoelectric control etc.. Moreover, CCA based features can be advantageous to capture the underlying statistics of the multiple views of EEG signal. Fig.4 shows the correlation of the variables for three subject groups. It is seen that only a few pair of vectors are highly correlated.



(a)



(b)



(c)

Fig3: Left: Approximation coefficients (A2s), when the DWT is applied on MTSs with prototype db2 wavelet from three typical EEG pattern, A (a), B (b), and E (c) subject. Right: Details (D2s) of respective MTSs.

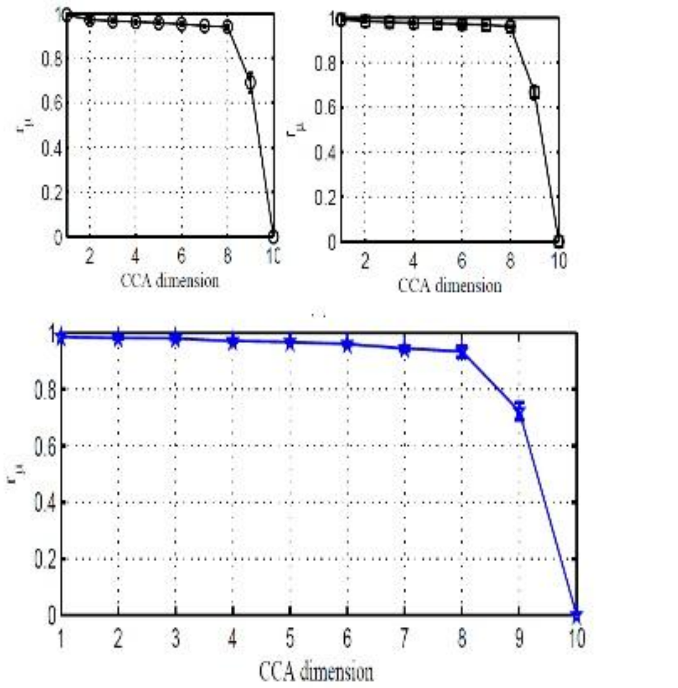


Fig. 4: Correlation of MTSs and its respective delay versions; a) B; b) A; c) E Each MTS is formulated by taking k=10 templates from each subjects. Horizontal axis represents the dimension of MTSs.

Such pairs of vectors are widely used in many applications. These vectors have emerged as useful means of analyzing single modality data or multimodality data or non-stationary signal analysis such as EEG. Thereby, highly correlated statistical pairs are used for feature fusions and exploit for supervised learning. It is well known that suitable choice of transform domain feature on fusion can contain more classification related information than direct signal domain features. Multi-view transform features, in essence, attempt to capture low dimensional classification related information. As a result, one can anticipate that compact features can be useful to discriminate various classes of EEG. The efficacy of proposed feature and fusion used herein is confirmed by graphical and statistical analysis as well. All these factors motivated the use of transform features in this work.

In this section, we describe feature reduction and fusion framework and finally evaluate discriminant patterns. Estimated highly correlated vectors are used to derive compact measures of features. Albeit its wide-spread use in image or character recognition, such measures are have only one recently been used in physiological signal analysis [23, 48]. In this work we extract d-dimensional transform vectors from each pair measurement and finally they are applied in feature fusion models. Assume that f(m)=f(A,B) is an d-dimensional mCCA feature function where m=1,2,3, and f(m)=f(C,D) is the DWT+ mCCA transform feature function. Both set of feature functions are statistically independent. Feature fusion is performed either by concatenation or summation of the transformed feature vectors. The vectors are as follows

$$z1 = \begin{pmatrix} X1 \\ X2 \end{pmatrix} = \begin{pmatrix} A^T & X_1 \\ B^T & X_2 \end{pmatrix} = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}^T \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

$$Z_2 = X_1^* + X_2^* = A^T X_1 + B^T X_2 = \begin{pmatrix} A^T \\ B^T \end{pmatrix}^T \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \quad (6)$$

Where $X_1^* = A^T X_1$, $X_2^* = B^T X_2$ are mCCA transformation and Z1 and Z2 are called the Canonical Correlation Discriminant Features (CCDFs). Further to incorporate orthogonal sets of CCDFs are generalized as Z11 = diag[A; ·;D][X1;……;X22] and Z22 = [A;…..;D][X1;…;X22].

This multi-domain feature fusion can be regarded as a multi-modal fusion. This way we evaluate all CCDFs and subsequently means are used as global feature descriptor for learning. As a matter of fact such combination of fusion has must robustness and thereby considered as good global descriptor. Use of such feature fusion strategy in our proposed method is inspired from the success of multi group feature fusion using multiset CCA [21, 22]. Unlike cited methods our approach derives global descriptors utilizing two orthogonal domain spaces and thereby could provide more discriminant power which has been explored in later. Besides, it is believed to be more suitable for classification related large scale problem as it includes multiple views' correlation information through compact distribution, i.e., it preserve as much as information so that model can easily identify any unknown distribution. Various limitations of previous methods motivate its use in the classification scheme propounded herein. In most decision models feature biasing is the major concern. It is mainly due to non-linear and non-stationary of signals and diversity in data populations. As a consequence in spite of being innate and simple it obscures many approaches unsuitable for real-life execution. Further some approaches meet the overall objective of model but involvement of multiple stages creates a bridge in and out-profession. It has been reported that CCA based feature extraction scheme is simple to implement and yields consistently good performance in pattern classification and other applications as well. Besides, superior performance of our method, as compared to conventional approaches also corroborates with this fact.

5.2 Efficacy of selected space and statistical validation

mCCA allows to derive compact features by reducing redundancy from large feature views, which have must discriminant ability in real-time applications [15]. Besides, it is conceptually simple, computationally inexpensive and efficient implementation is feasible. Again correlation between two set of vectors indicates extent of similarities. Accordingly the nature of variability and quantitative measures can be specified. Major concern in such learning strategy is the ignorance of class structure among the samples. Nonetheless, in classification problems we are interested in effective separation of classes with reduced sets of features. In such case two stage approach CCA+LDA will be effective solution. This preserves pair-wise correlation measures and classes information within each set of features as well. Superior performance in multimodal recognition problem motivates the use of this scheme in our study [5]. However they omit efficacy of feature patterns. On the other, our method can extract and fuse large scale information for effective learning and it is free from limitations and therefore, it is very suitable for EMG signal staging. Further, the model can efficiently combine statistically independent features associated with same input elicited via fusion strategy. LDA feature projection technique finds linear transformation that maximizes between-class scatter matrix and minimizes the within class scatter [17]. It seeks coordinate system to enhance the separation margin among various groups of projected feature. A linear transformation is expressed as $y = W^T x$, where x is the original feature vector with n dimensionality, y is the projected feature vector with k dimensionality, and W is an n X k matrix. The LDA is mainly based on family of two scatter matrices. The within-scatter matrix S_W and between-scatter matrix S_B can be defined as-

$$S_W = \sum_c N_c (\mu_c - \mu)(\mu_c - \mu)^T \quad (7)$$

$$S_B = \sum_c (x - \mu_c)(x - \mu_c)^T$$

Where μc, μ and Nc stand for class mean, overall mean of the entire sample set and number of samples in class C respectively. Finally, total scatter matrix is ST = SW + SB. For a scatter matrix the measure of spread is the determinate. Thus matrix W is identified that maximizes the learning criteria

$$\frac{W^T S_B W}{W^T S_W W}$$

The matrix W composed of k eigenvectors corresponding to the k largest Eigen values of S_W^{-1} SB. Since maximum rank of matrix SB is K-1, the value of k must be defined as less than K. Therefore number of classes limits the dimension of projected space. To avoid this limitation the total scatter matrix is used instead of the between-class scatter matrix in the learning criterion [17]. From this analysis k dimensional feature y-obtained. In this analysis after projection, we extracted combined fused 5-dimensional feature matrices, which are further subjected to statistical hypothesis test. This filtered-based feature selection tools evaluates the efficacy of selected feature space and make confirm whether selected features are statistically significant or not [46]. Further, it reduces the feature space by removing the insignificant feature from the selected space. Thus it gives vibrant image of model performance even before assessing the performance of models. Here to ensure weather selected features have the discriminate capability among the various classes; we execute a one-way analysis of variance (ANOVA) [46]. The test is carried out in MATLAB at 95% confident level. Thus, any feature having $p > \alpha (= 0.05)$ is considered as insignificant. It thus too substantiate with our aforementioned policy for logical conclusion. In other words, statistical fused features deployed in our method have good discriminant ability. Thus, it can be expected good algorithm performance while fed them into classifiers and it is efficient policy to classify three groups of studied subjects.

5.3 Performance Comparison

The performance of the proposed feature level fusion algorithm is compared with that of several state-of-the-art feature level, matching score level and decision level fusion algorithms.(Table.IV). EEG signals (A, B and E) were categorized using DWT-based statistical model, k-nearest neighbour (KNN) [37], Artificial Neural Network (ANN) [38], Nueoro-fuzzy system (ANFIS) [39], and Quadratic approach [40]. Nonetheless, use of DWT requires human interference and sub-band statistics reduces the curse of dimensionality for wide inconsistency of templates, however, it may not be feasible for the analysis. Further, inherent non-stationary nature of templates, use of DWT features lead to result diversity. Apart from reported works, this study is based on subspace learning, through which promising results are observed in comparison to results reported in the literature. It is seen that the proposed algorithm misclassified one case (A), which presumably due to inherent similarity of templates and the order selection before fusion or the extracted features are significantly different from trained pattern. The integrity of the proposed technique is due to well-defined feature extraction and fusion strategy which in turn decreases the complexity. Further, unique features can be easily captured, which plays a crucial role in multi-tasking learning models. Additionally, the algorithm involves a number of steps and it takes mean execution time 89.72 s approximately over three runs. However, it could vary with machine.

Table 1. Comparison of proposed one with the State of Arts Techniques.

Classifier	**Feature type**	**Study group/Object**	**Overall Accuracy in %**
KNN	DWT	5/10	97
ME, ANN	DWT	2/10	94.50/93.20
ANFIS	DWT	5/10	92.55
QUADRATIC	DWT-FFT	3/10	98.70
PROPOSED	**DWT+CCA+SF***	**3/10**	**99.33**

* Statistical Features

5.4 Discussions and conclusions

This article adopts a feature fusion based classification scheme for single channel based automatic staging of EEG. Apart from conservative approaches combined hybrid models are introduced for effective classification of various stage of EEG. Our study used technique to extract combined features from multiple views generated from signal, whereas some of prior studies such as in [41, 42] extracts feature (DWT or statistical measures) from specific set of signals to classify entire data set. Further, due to wide variety of data it is difficult to identify unknown class distribution based on the learned patterns derived from them. Taking specific frames of signal such as EMG or EEG for feature extraction can be major source of error due to their non-linear and non-stationary behaviour. Although significant level of performance of such data-driven approaches with specific set of data, it may fail maintaining the consistency with wide variety of data sets. Besides involvement of multiple stages for feature extraction can also be potential source of error. It is seen that many prior works derive class distribution using various techniques and in case of large dimensionality of feature statistical measures such as Mean, Standard Deviation etc. are deployed as feature to the models. In case of close data distribution of various classes such parameters may not represent meaningful measures. Additionally, assumptions based models are usually complicated to understand and requires maintaining multiple constraints. The integrity of such model not easily understandable for common reader and thereby, it obviates practical implementations. Furthermore, such model works well in one domain but fails in other [43]. This is apparently due to assumptions, of which some are unrealistic in nature [26]. The proposed scheme in this work therefore does not have the aforementioned limitations. Fourier transform is a popular technique of signal analysis; however it assumes the signal to be linear and stationary whereas bio-medical signals such as EMG, EEG etc. signal are known be highly nonlinear and non-stationary. Multi-resolution wavelet transformation based methods have been employed successfully in EMG classification task [44,47,49]. In wavelet transformation the fixed frequency scale depends on sampling frequency and level of decomposition. Besides, appropriate choice of basis function that to be used in the transformation is essential for effective model performance. Domain subjective knowledge and prior works help to select order of decomposition and basis function. It is worth mentioning the fact that in the proposed method we apply DWT on feature data set only to create statistically independent views. As both domains correspond to same input pattern and thereby, fused feature space evaluated from independent spaces could lead to better discriminate ability of the models. Introduction of this combined approach is motivated from multi-modal fusion application as in [13] and multi-view learning [2]. Our proposed mCCA does not involve any assumption and complicated steps and is entirely data-driven, which make it attractive choice for processing and analysis of biomedical signals. Further, proposed feature fusion based approaches yield higher discriminate ability than commonest approaches. Thus, it cogently manifests the efficacy of fused features. It was our principal motivation for choosing mCCA based data-driven approach over other conventional techniques. Additionally, it controls the complexity, feature biasing and over-fitting, common weakness of traditionally used learning models. These factors in conjunction with afore mentioned advantages have pushed toward attaining higher accuracy of mCCA based learning. For real-world applications, implementation of the proposed algorithm can be extended to develop user-friendly graphical user interference (GUI) and it makes the whole analysis in a systematic way to speed up the process even further. This analysis is however carried out over EEG signal analysis; it can be implemented to solve other non-stationary signals analysis such as ECG, EMG and various pathological state detection problems. How the algorithm behaves with other efficient boosting algorithm can be potential topic of further research. We thus conclude, as the algorithm aftermaths evince that our method for automatic staging of EMG is effective and efficient. Thus, it promotes large scale population research and push toward for portable device implementation.

6. ACKNOWLEDGEMENT

The authors would like to acknowledge Computer Laboratory and Advanced Level Biotech Hub, Pub Kamrup College, Baihata Chariali, Kamrup, Assam, India for availing the facility.

REFERENCES

[1] Xia Xia, T., Tao, D., Mei, T., et al.: 'Multiview spectral embedding', IEEE Trans. Syst. Man, Cybern. B, Cybern.2010, 40 (6), pp.1436–1446

[2] Liu Liu, M., Zhang, D., Shen, D., et al.: 'Inherent Structure Based Multi-view Learning with Multi-template Feature Representation for Alzheimer's Disease Diagnosis', IEEE Trans. Biomed.Eng.,2015 (in press)

[3] LH Lahat, D., Adali, T., Jutten, C.: 'Multimodal data fusion: an overview of methods, challenges, and prospects', Proceedings of the IEEE, 2015, 103(9), pp.1449–1477

[4] Bing Sun, B. Y., Zhang, X. M., Li, J., et al.: 'Feature fusion using locally linear embedding for classification', IEEE Trans. Neural Network., 2010, 21 (1), pp. 163–168

[5] MH Haghighat, M., et al.: 'Discriminant correlation analysis: Real-time feature level fusion for multimodal biometric recognition', IEEE Trans. Inform. Forensics Security, 2016, 11, pp.1984–1996

[6] R2 Zhou, D., Burges, C.: 'Spectral clustering and transductive learning with multiple views'. Proc. 24th Int. Conf. Mach. Learn., Corvallis, Jun. 2007, pp. 1159–1166

[7] ARR1 Hassan, A. R., Haque, M. A.: 'Computer-aided gastrointestinallhemorrhage detection in wireless capsule endoscopy videos', Computer methods and programs in biomed., 2015, 122 (3), pp.341–353

[8] ARR2 Hassan, A. R., Subasi, A.: 'Automatic identification of epileptic seizures from EEG signals using linear programming boosting', computer methods and programs in biomed., 2016, 136,pp. 65–77

[9] Pachori, R. B., Bajaj, Y.; 'Analysis of normal and epileptic seizure EEG signals using empirical mode decomposition', Computer Methods and Programs in Biomed., 2011, 104 (3), pp.373–381

[12] She Shen, X., Sun, Q.: 'A novel semi-supervised canonical correlation analysis and extensions for multi-view dimensionality reduction', J. Vis. Commun. and Image Represent., 2014, 25 (8),pp. 1894–1904

[13] Sargin, M. E., YÃijcel, Y., Engin, E.: 'Audiovisual synchronization and fusion using canonical correlation analysis', IEEE Trans. Multimedia, 2007, 9 (7), pp. 1396–1403

[14] Jun1 Chu, J. C., Moon, I., Mun, M. S.: 'A real-time EMG pattern recognition system based on linear-nonlinear feature projection for a multifunction myoelectric hand', IEEE Trans. Biomed. Eng.,2006, 53 (11), pp. 2232–2239

[15] Fun Fu, Y., Cao, L., Guo, G., et al.: 'Multiple feature fusion by subspace learning Proceedings of the 2008 international conference on Content-based image and video retrieval'. Proc. ACM int. conf. Content-based image and video retrieval, 2008, New York, NY, USA

[16] Dutta, L., Hazarika, A., Bhuyan, M.: 'Comparison of Direct Interfacing and ADC Based System for Gas Identification using E-Nose'. IEEE conf. on Inventive Computation Technologies,2016, India

[17] Jun2 Chu, J. C., Moon, I., Mun, M. S., et al.: 'A supervised feature-projection-based real-time EMG pattern recognition for multifunction myoelectric hand control', IEEE/ASME Trans.b Mechatronics, 2007, 12 (3), pp. 282–290

[18] Dutta, L., Hazarika, A., Bhuyan, M.: 'Microcontroller Based E-Nose for Gas Classification without Using ADC', Sensors and Transducers, 2016, 202 (7), pp. 38–45

[19] Kim, T. K., Kittler, J., Cipolla, R.: Discriminative learning and recognition of image set classes using canonical correlations', IEEE Trans. Pattern Anal. Mach. Intell., 2007, 29 (6), pp. 1005-1018

[20] Hotelling, H: 'Relations between two sets of variates', Biometrika, 1936, 28 (3/4), pp. 321–377

[23] Khushaba, R. N.: 'Correlation analysis of electromyogram signals for multiuser myoelectric interfaces', IEEE Trans. Neural Syst. Rehabil. Eng., 22 (4), pp. 745-755

[24] Sun, L., Shuiwang, J., Jieping, Y.: 'Canonical correlation analysis for multi label classification: A least-squares formulation, extensions, and analysis', IEEE Trans. Pattern Anal. Mach. Intell.,33 (1), pp. 194–200

[25] DeClercq,W., Vergult, A., et al.: 'Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram', IEEE Trans. Biomed. Eng., 2006, 53 (12), pp.2583-2587

[26] Correa, N. M., Adali, T., et al.: ’Canonical correlation analysis for feature-based fusion of biomedical imaging modalities and its application to detection of associative networks inschizophrenia’, IEEE J. Sel. Topics Signal Process., 2008, 2 (6), pp.998–1007

[27] H. Hassanpour, M. Mesbah, B. Boashash, “Time-frequency feature extraction of newborn EEG seizure using SVD-based techniques,” EURASIP J. on Applied Signal Proces., vol. 2004, pp. 2544-2554, 2004.

[28] A. Subasi, M. I. Gursoy, “EEG signal classification using PCA, ICA,LDA and support vector machines,” Expert Syst. Appl., vol. 37, no. 12,pp. 8659-8666, 2010.

[29] Mahyar, Z. K., Bruce, W. C., et al.: ’EMG feature evaluation for movement control of upper extremity prostheses’, IEEE Trans. Rehabil. Eng., 1995, 3 (4), pp. 324âˇA ¸S-333

[30] A. R. Hassan, A. Subasi, “Automatic identification of epileptic seizures from EEG signals using linear programming boosting,”Comput. Methods Programs Biomed., vol. 136, pp. 65-77, 2016.

[31] A. R. Hassan, M. A. Haque, “Computer-aided obstructive sleep apnea screening from single-lead electrocardiogram using statistical and spectral features and bootstrap aggregating,” Biocybern. Biomed., vol. 36, no. 1, pp. 256-266, 2016.

[32] A. R. Hassan, “Computer-aided obstructive sleep apnea detection using normal inverse Gaussian parameters and adaptive boosting” Biomed Signal Proces. Control, no. 29, pp. 22-30, 2016.

[33] A. R. Hassan, M. I. H. Bhuyan, “Automatic sleep scoring using statistical features in the EMD domain and ensemble methods,”Biocybern. Biomed., no. 36, no. 1, pp. 248-255, 2016.

[34] U. M. H. Orhan, M. Ozer, “EEG signals classification using the K-means clustering and a multilayer perceptron neural network model,” Expert Syst. Appl., vol. 38, no. 10, pp. 13475-13481, 2011.

[35] M. H. Soomro, S. H. A. Musavi, B. Pandey, “Canonical Correlation Analysis and Neural Network (CCA-NN) Based Method to Detect Epileptic Seizures from EEG Signals,”Int. Journal of Bio-Science andBio-Tech., vol. 8, no. 4, pp. 11-20, 2016.

[36] M. K. Kiymik, M. Akin, A. Subasi,“Automatic recognition of alertness level by using wavelet transform and artificial neural network,” Journal of neuroscience methods, vol. 139, no. 2, pp. 231-240, 2004.

[37] U. Orhan, M. Hekim, M. Ozer, “EEG signals classification usingthe K-means clustering and a multilayer perceptron neural network model,”Expert Syst Appl, vol. 38, pp. 13475, 2011.

[38] A. Subasi, “EEG signal classification using wavelet feature extraction and a mixture of expert model,” Expert Syst Appl, vol. 32, pp. 1084,2007.

[39] M. Hekim, “The classification of EEG signals using discretization-based entropy and the adaptive neuro-fuzzy inference system,” Turk J. Elec.Eng. Comp. Sci., vol. 24, pp. 285- 297, 2016.

[40] Gajic D. et al., “Detection of epileptiform activity in EEG signals based on time-frequency and non-linear analysis,” Frontiers in Comput. Neurosci., vol. 9, pp. 38, 2015.

[41] S2 Subasi, A.:’ Medical decision support system for diagnosis of neuromuscular disorders using DWT and fuzzy support vector machines’, Comput. Biol. Med., 2012, 42 (8), pp. 806–815

[42] S1 Subasi, A.: ’Classification of EMG signals using PSO optimized SVM for diagnosis of neuromuscular disorders’, Comput. biol. med., 2013, 43 (5), pp. 576–586

[43] A1 Barthakur, M., et al.: ’Rule based fuzzy approach for peripheral motor neuropathy (PMN) diagnosis based on NCS data’. proc. IEEE int. Conf. Proc. Recent Advances and Innovations in Eng., 2014, India, pp. 1-9

[44] S3 Subasi, A. “Classification of EMG signals using combined features and soft computing techniques”, Appl. soft comput., 2012,12 (8), pp. 2188–2198

[45] Doulah, A. B. M. S. U et al.:’Waelet Domain Feature Extraction Scheme Based on Dominant Motor Unit Action Potential of EMG Signal for Neuromuscular Disease Classification’, IEEE Trans. Biomed. Circuits Syst., 8 (2), pp. 155–164

[46] Nikolic, M.: ’Detailed Analysis of Clinical Electromyography Signals: EMG Decomposition, Findings and FiringPattern Analysis in Controls and Patients with Myopathy and Amytrophic Lateral Sclerosis’. PhD thesis, Faculty of Health Science, Univ. Copenhagen, Copenhagen, Denmark, 2001

[47] S7 Subasi, A.: ’A decision support system for diagnosis of neuro muscular disorders using evolutionary support vector machines’, Signal, Image Video Process., 2015, 9 (2), pp. 399–408

[48] S5 Gokgoz, E., Subasi, A.: ’Comparison of decision tree algorithms for EMG signal classification using DWT’, Biomed. Signal Process. Control, 2015, 18, pp. 138–144

[49] Ag Parsaei, H., Stashuk, D.W.: ’Adaptive motor unit potential train validation using MUP shape information’, Med. eng. physics,2011, 33 (5), pp. 581–589