Chapter 4

Feature fusion using multi-view discriminant correlation analysis

4.1 Introduction

For complete knowledge of the phenomenon of interest, thorough analysis of availability of multiple responses is mandatory requirement. As mentioned in Ch.3, analysis of such high dimensional multi-view associated with high impact-biomedical, commercial, social, military and environmental applications, promotes deriving efficient methodologies [1]. At present, such studies are of high demand and reach far beyond pure academic interest. The previous chapters have already addressed a mathematical framework incorporating multiple domains multi-view feature fusions with an aid of CCA and mCCA. With wide prospective on multi-class learning, this chapter aims at establishing a novel multi-domain multi-view DCA (mmDCA) fusion based data-driven model generalizing DCA, which is inherently capable of incorporating the class association in the feature space unlike CCA or mCCA. DCA contains similar theoretical formulation with the CCA with exception of feature discriminant function [6]. mmDCA extends the theory of DCA to find the multiple domain canonical variates summarizing the correlation structures among the multiple input features by linear transformation. In contrast to the DCA where correlation between pair feature is maximized and well separated, it, in addition, optimizes the criterion function between features in multiple domain independently to obtain maximum overall correlation and more meaningful structure corresponding to same phenomena.

The adopted model eliminates the between-class correlations and restricts the correlations within the classes. Additionally, it has the ability to decorrelate feature vectors which are common to different classes within each feature sets. Such design focus in the early phase of learning helps mobilizing the performance rather than the conventional multi-step approaches that require tuning or standardization of parameters

to meet the objective. In addition, the adopted approach does not encounter small-sample size problems that usually occur in many real-world problems. To the best of author's knowledge, no such approach has been put forwarded in literature. The chapter identifies many shortcomings in current approaches like class-structure or class association, complexity, performance, reliability for real-time implementation, adaption to noise and signal processing method.

The elaborated well-defined framework is to gain benefit over wide-scale learning scenario. Extensive experimental investigations on EMG data demonstrate the effectiveness of adopted method that outperforms many state-of-the-art methods. In addition to previously mentioned issues, the proposed model aims at the best achievable error rate, interpretability, uniqueness and requirement of straightforward approaches.

4.2 Key ingredients

As stated earlier, the CCA does not take into account the orthogonal components which may play a dominating role in pattern recognition tasks. As a remedial measures, our feature fusion approach via DWT+CCA aims to synchronize domain independent features with directly evaluated features in order to maintain the integrity of good learning approach. The proposed mmDCA scheme is implemented to extract a sets of local features, which will be termed as local generalized CCDFs and to fuse all discriminant evaluations in a synchronized manner after incorporation of class-specific information into the corresponding feature sets. It is believed that the mmDCA will become more reliable and it can overcome the shortcoming of the previously described strategies.

To avoid computational and theoretical bottlenecks, many generalized versions such as MCCA and method in [34, 139]. However, these methods do not expose the integral relationship among the multi-set variables and the constraints do not guarantee that the transformed variables are statistically uncorrelated [30]. MICCA [29] precisely expresses the integral correlation among multi-set features. Nonetheless, it pursues an iterative approach that reduces its efficiency. Most recently, sparse representation has garnered interest for both reconstructive and discriminative tasks [140–142]. The main assumption of these models is that a query sample belonging to a specific class can be viewed with a linear combination of the training samples from that class. Thereby, it searches for a sparse vector having non-zero elements only in the indices corresponding to that class. It does not follow the principle of feature fusion that requires formulating single feature vector to be utilized in any classification model as mentioned in [143]. Nonetheless, Joint Sparse Representation classification is regarded as a fusion technique that creates multiple corresponding dictionaries each using training samples of a modality. Taking the query consisting of multiple modalities, it attempts to evaluate joint sparse vectors that share the same sparsity pattern and have non-zero values only in the

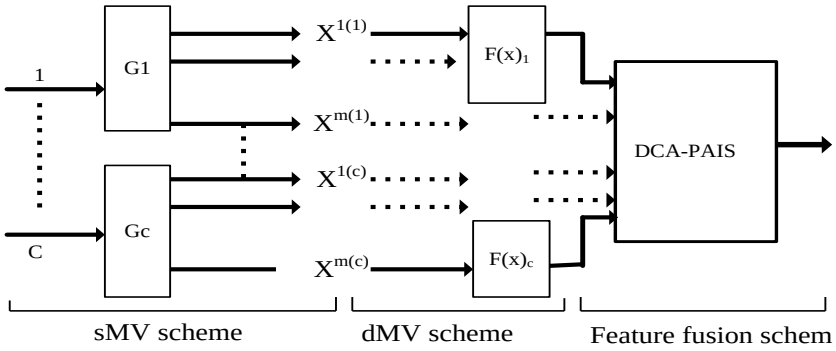

Fig. 4-1: Block diagram of proposed scheme for input process measrement. Here $F(x)_c$ represents function of dMV features generated from sMV features corresponding to m.

indices corresponding to a mutual class in multiple modalities. That is, it utilizes training samples of the same class from the different modalities to reconstruct the query data. The authors in [144] introduce a multimodal task-driven dictionary learning algorithm that improves the performance of this approach.

Although previous chapters have already addressed the methodologies of MVL and exemplified, in order to further meliorate algorithmic performance incorporating multiple factors, a new model, namely, mmDCA is addressed with comprehensive details.

4.3 Multi-view feature generation scheme

In addition to the previously described schematic representation of MV learning network (Ch.3, Fig. 3-5), in the same vein, a more generalized mathematical framework is important and possibly required for better concrete meaning. The proposed model as shown in Fig. 4-1 tediously examines the intelligent integration of information that evolves underlying phenomena and advanced inference logic. As is evident in the previous chapter (Ch. 3, Section 3.4), the multi-view generations are based on nature of subjective signals, e.g., age of subjects, disease profiles (i.e., mild, severe), duration and instrumental setting. Taking a dominating factor like age that influence the distribution of data patterns in signals, the previous experiment have already argued for an efficient support framework and accordingly it achieved the goal. Despite the achievements, this chapter discusses another approach using robust edition of CCA that includes class information for a true interaction of the signals inherently associated with subject groups at the ground level.

4.3.1 Multi-view direct feature

This section describes feature extraction scheme similar to the Fig. 3-7 of previous chapter. A given dataset includes C input processes and each process contains signals which

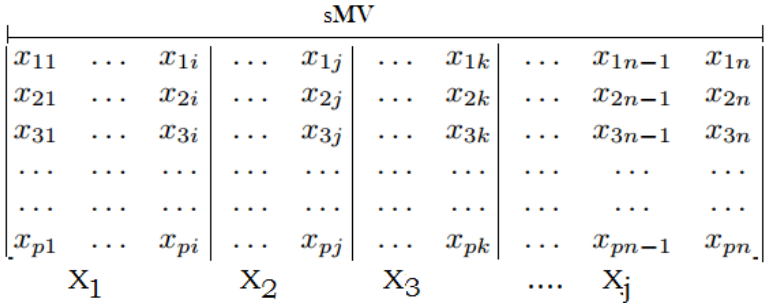

Fig. 4-2: Diagram shows feature generation scheme by segmenting the high dimensional feature matrix sMV$\in \mathbf{R}^{p\times n}$ formulated from a relevant set of signals. Each row represents 1D EMG signal of 11.2 s duration. Here, $X_1, \ldots, X_i$ indicate equally decomposed multiview features formulated using undertaking strategy. p and n represent number of signals and samples in each signals.

Table 4.1: Categorical partition of studied subject groups and sMV which are function of dMV features. Note that $j = 1, 2$ represents two domains raw+mmDCA and raw+DWT+mmDCA.

Category [C]	Groups [G]	Age range [Year]	Sub-views [Decomposition]
ALS	$G_1(X^{1(1)}, \ldots, X^{m(1)})\forall j$	35-52,53-61,62-67	$F(x)_1 = [X_1, \ldots, X_i]\forall j$
Myopathy	$G_2(X^{2(1)}, \ldots, X^{2(c)})\forall j$	19-26,28-41,44-63	$F(x)_2 = [X_1, \ldots, X_i]\forall j$
Normal	$G_c(X^{1(c)}, \ldots, X^{m(c)})\forall j$	21-26,27-29,29-37	$F(x)_c = [X_1, \ldots, X_i]\forall j$

are recorded at same or different experimental conditions. To ensure more diversity in dataset entry, each process is partitioned into G_C subgroup depending on a given condition which varies with the process. From each G_C, m feature matrices $X^{m(C)}$, referred to as sub-multi-view (sMV) are evaluated and it is extended to all processes as shown in Fig. 4-1. The sMV is further decomposed uniformly to evaluate decomposed-MV (dMV) feature as shown in Fig. 4-2, where each row indicates 1D-signal with samples x_{pn}. The reason of decomposition is to find inherent mutual information between two consecutive features using CCA or DCA while dissimilar components due to noise sources are eliminated.

4.3.2 Multi-view DWT feature

Recalling the Section 3.3.3, Ch. 3 that explains the use of second level DWT with *db*2 [58], it aims to use localized low-frequency components for deriving the features since most energy contents of biomedical signals, specifically EMG and EEG, usually fall in low frequency range, e.g., 0-1 kHz for EMG [16]. The higher level of decomposition further narrow downs the localization of frequency in subbands coefficients. The low frequency a^2 components of all selected signals are used as input signals to find DWT-sMV and dMV similar to section 4.3.1.

As mentioned earlier, the hypothesis is, therefore, to create a set of MV or dMV by employing multiple signals through a specific strategy which could have most physiologically relevant information and energy contents. The mathematical formulation

used herein is only a concatenation of a more general idea that applied to data sets. Nonetheless, it is worth mentioning that the decomposed vectors, $X_1 \in \mathbb{R}^{q \times p}$, $X_2 \in \mathbb{R}^{q \times p}$ with p samples each, do not necessary to have same dimensional vectors. It is only for ease of implementation and to maintain the integrity of our approach (note that $P = \sum_i p = n$ samples). It is natural to incorporate dimension reduction strategy in case of large dimensional input features before mmDCA. However, it does not lose the generalization ability of input space, in fact, it controls the complexity and ascends the learning ability curtailing the parameters setting.

4.4 Feature fusion using mmDCA

Recalling the within-sets covariance matrices Σ_{11} and Σ_{22} and between-set covariance matrix of views Σ_{12} are of dimension $q \times q^*$. The overall covariance matrix Σ containing all information on association between pairs of variables is as below:

$$\Sigma = \begin{bmatrix} Cov(X_1) & Cov(X_1, X_2) \\ Cov(X_2, X_1) & Cov(X_2) \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \tag{4.1}$$

It is to be noted that in case of inconsistent pattern vectors, the understanding of the relationships between these two sets of feature vectors from this matrix is quite difficult task [145]. CCA handles this issue efficiently by subspace transformation as described in the previous Chapter (Ch. 3, Section 3.3.1), which finds sets of transformation and corresponding eigenvalue matrix α^2 with $d = rank(X_1, X_2) \leq min(p, q)$ non-zero elements with descending ordered correlations. Here p and q represent number of rows and columns in the feature matrices. With an appropriate choice of dimensionality, the subspace features are fused using Eq.(3.13b)-(3.14a) to find the gCCDFs as comprehensive statistics of underlying information of pair features. In case of large number of MVs or dMVs, then the gCCDFs evaluating from all possible pairs of features are necessary with the help of the Eq.(3.14b)-(3.14c). For the case of low deviation among the gCCDFs stored in DMI and DMII, second level fusion, i.e., F=2 as defined in the section 3.3.4, the global mean gCCDFs is more suitable in context of complexity and dimensionality concerns. If the order choice of d in evaluating the gCCDFs is low, then the DMI and DMII, F=2 provide more concrete meaning. Even in case of large dimensionality of DMI and DMII, feature reduction techniques help transforming them to low dimensional vectors. The integrity of adopted CCA-based fusion schemes is made clear to users by demonstrating the performance in previous studies. However, some important issues for the sake of being well multi-classification task are needed to be explored further.

Singularity and ignorance of class information are two disputable issues in CCA-based learnings. Singularity issue can be handled efficiently by introducing regulariza-

*$\Sigma_{12} = \Sigma_{21}^T = X_1^T X_2$, Σ_{11}, Σ_{22}, Σ_{12} and Σ_{21} are presented as Σ_{xx}, Σ_{yy}, Σ_{xy} and Σ_{yx} according to the input feature vectors X and Y in chapter 3, Section 3.3.1

tion or combination of both PCA and CCA/mCCA, i.e., PCA+CCA/mCCA [**?**]. Based on class-information, the dimension reduction techniques, also known as visualized tools such as LDA, SOFM, determine an effective feature distribution in decision surface with enhanced class separation margin. In case of feature vectors obtained from CCA space for various objects may have the possibility of similarities among different class specific feature vectors that causes overlapping of feature clusters and degradation of model performance.

In this contrast, in a two-stage approach LDA + CCA/mCCA, CCA/mCCA fails to preserve the properties achieved in LDA+CCA. However, the reverse case CCA/mCCA+LDA copes the class structure efficiently (Ch. 3, Section 3.5.2.1). In such case, LDA work only over the information preserved by CCA or mCCA. Unlike CCA/mCCA and DCA, the advocated model simultaneously makes the correlation among features and decorrelation between different group features in multi-domains (i.e., $j = 1, 2$). Thus, it could fulfill the requirement for efficient design module. The combination of correlation analysis and discriminant analysis has also been utilized in earlier work [21] and [146]. However, problem definition and formulation are different from the adopted approach. For instance, Kim *et al.* [21] addresses cross-view face recognition systems by using correlation analysis, wherein it aims to find the correlated features from feature vectors of the various domains.

4.4.1 Preliminaries and theoretical approach

Given a data matrix X with n columns is composed of samples collected from C separate classes of dataset and n_i columns corresponding to i^{th} class. Sample and global mean of feature matrix are $\bar{x_i}$ and μ respectively. To separate the class within the feature set, it is required to maximize the between-scatter matrix S_B.

$$S_{Bx1} = \sum_{c} n_i(\nabla_i)(\nabla_i)^T = \Phi_{bx}\Phi_{bx}^T, \tag{4.2a}$$

$$\Phi_{bx(p\times C)} = [\sqrt{n_1}(\bar{x_i} - \mu), \ldots, \sqrt{n_C}(\bar{x_i} - \mu)]. \tag{4.2b}$$

where[†] $\nabla_i = \bar{x_i} - \mu$. In case of large feature vectors in comparison to the classes C, the solution to the RHS of Eq.(4.2a) can be evaluated using mapping technique which involves $C \times C$ covariance matrix [6]. To ensure well-class separation, the symmetric positive semi-definite matrix $\Phi_{bx1}^T\Phi_{bx1}$ is diagonalized as-

$$P^T(\Phi_{bx1}^T\Phi_{bx1})P = \bar{\Lambda} \tag{4.3}$$

where P is the orthogonal eigenvector and $\bar{\Lambda}$ is the diagonal matrix having real non-negative eigenvalues in descending order. Let $Q_{C\times r}$ matrix contain first r non-zero

[†]where $\bar{x_i} = \frac{1}{n_i}\sum_{j=1}^{n}{}_i x_{ij}$ and $\mu = \frac{1}{n}\sum_{i=1}^{c} n_i\bar{x_i}$

eigenvectors which correspond to the largest eigenvectors of P.

$$Q^T(\Phi_{bx1}^T\Phi_{bx1})Q = \Lambda_{r\times r} \tag{4.4}$$

The most significant eigenvectors of S_{Bx1} can be found by mapping: $Q \longrightarrow \Phi_{bx1}Q$. That is $(\Phi_{bx1}Q)^T S_{Bx1}(\Phi_{bx1}Q) = \Lambda_{r\times r}$. $W_{bx1} = \Phi_{bx1}Q\Lambda^{1/2}$ transformation utilizes S_{Bx1} to reduce X_1 from p to r, i.e.,

$$W_{bx1}^T S_{Bx1} W_{bx1} = I, \tag{4.5a}$$

$$\acute{X_1}_{r\times n} = W_{bx1(r\times n)}^T X_{1(p\times n)}. \tag{4.5b}$$

$\acute{X_1}$ is the transformed pattern of X_1, where classes are well separated. There are at most $C-1$ nonzero eigen vectors. Following the same principle, second feature variable X_2 is mapped to $\acute{X_2}$, where transformed matrix W_{bx2} uses the between-class scatter matrix for the second modality profile S_{Bx2} and minimizes dimension from q to r (see Eq.(4.6a)-(4.6b)).

$$W_{bx2}^T S_{Bx2} W_{bx2} = I, \tag{4.6a}$$

$$\acute{X_2}_{r\times n} = W_{bx2(r\times q)}^T X_{2(q\times n)}. \tag{4.6b}$$

The modified version $\Phi_{bx1}^{'}$ and $\Phi_{bx2}^{'}$ of dimension $r \times C$ are orthogonal matrices and ${\Phi_{bx1}^{'}}^T\Phi_{bx1}^{'}$ and ${\Phi_{bx2}^{'}}^T\Phi_{bx2}^{'}$ are diagonally dominating in nature, i.e., if a_{ij} are the matrix elements, then $\forall i, |a_{ii}| > \sum_{i\neq j}|a_{ij}|$. The non-diagonal elements are close to zero, i.e., $a_{ij} \longrightarrow 0, i \neq j$ and diagonal elements tends to one, i.e., $a_{ii} \longrightarrow 1$. In CCA, it is necessary to ensure that feature vectors in one set have nonzero correlation only with their corresponding feature vectors in the other set for which the between-set covariance matrix ($\bar{\Sigma_{12}}$) of the mapped feature vector sets is diagonalized using SVD.

$$\bar{\Sigma_{12(r\times r)}} = U\Sigma V^T \Rightarrow U^T\bar{\Sigma_{12}}V = \Sigma. \tag{4.7}$$

Σ is a diagonal matrix. It is to be mentioned that $rank(\tilde{X_1}, \tilde{X_2}) = r$ and $\bar{\Sigma_{12(r\times r)}}$ form non-degenerate pattern[‡]. Let $W_{cx1} = U\Sigma^{-1/2}$ and $W_{cx2} = V\Sigma^{-1/2}$ so that $W_{cx1}^T\bar{\Sigma_{12}}W_{cx2} = I$ which utilizes the between-set covariance matrix. Finally transformation for X_1 and X_2 are as follows:

$$X_1^\star = W_{cx1}^T\bar{X_1} = \underbrace{W_{cx1}^T W_{bx}^T} X_1 = W_{x1}X_1, \tag{4.8a}$$

$$X_2^\star = W_{cx2}^T\bar{X_2} = \underbrace{W_{cx2}^T W_{bx2}^T} X_2 = W_{x2}X_2. \tag{4.8b}$$

where $W_{x1} = W_{cx1}^T W_{bx1}^T$ and $W_{x2} = W_{cx2}^T W_{bx2}^T$ signify the transformation of X_1 and X_2 respectively. Further, the between-class scatter matrix of transformed pattern is still diagonalized, i.e., $S_{Bx1}\star = W_{cx1}^T \underbrace{W_{bx1}^T S_{Bx1} W_{bx1}} W_{cx1} = \sum_{i,i} a_{ii}$. Now similar to CCA approach as mentioned in Chapter 3 (i.e, Ch. 3, section 3.3.4), the FVs are fused to

[‡] $\bar{\Sigma_{12(r\times r)}} = \bar{X_1}{\bar{X_2}}^T$

obtain discriminant features which can be expressed mathematically as:

$$\vartheta(X_1, \ldots, X_2^w) = \sum_{i=1}^{d} \sum_{i=1}^{d} \Theta(X_1, X_2) \oplus \zeta(X_1^w, X_2^w) \tag{4.9}$$

$$= \sum_{i=1}^{d} \sum_{i=1}^{d} \Theta(X_1^\star, X_2^\star) \oplus \zeta(X_1^{\star w}, X_2^{\star w}) \tag{4.10}$$

$$= \Pi_{i=1}^{d} \Delta(A, \ldots, D) \otimes \sigma(X_1, \ldots, X_2^w) \tag{4.11}$$

where $\Theta(\bullet)$ and $\zeta(\bullet)$ are statistically independent multi-domain feature functions, and $\oplus$ indicates fusion of features. Furthermore, $\Delta(.)$ and $\sigma(.)$ are two independent feature functions.

4.4.2 Proposed mmDCA-based learning

The proposed scheme incorporates all locally evaluate features using multi-domain DCA and to combine them into a single form of discriminant vector that could provide underlying phenomena associated with the formulated variables from the available data. This way, it handles a complex nonlinear problem into a simple correlation based problem. According to the formulation under discriminant framework, the mmDCA finds two sets of transformed features-one directly from the pair dMV of direct feature function $F(X) = f(X_1, \ldots, X_j)$ and other from the corresponding dMV in wavelet transformed space $F(W) = f(X_1(w), \ldots, X_j(w))$. Finally, mmDCA combines information through feature fusion strategy. Thus, explicit coupling of these information as shown in Eq.(4.11) without a prior assignment of individual is a cornerstone for preceding broad information towards logical inference to make the formulation natural, efficient and intuitive as possible.

Fig. 4-3 shows the block diagram of generalized scheme of mmDCA. It extends the theory of DCA to find the multiple domains canonical variates summarizing the correlation structures among the multiple variables by linear transformation. It includes the class-structure information to the feature space so as to enhance the discriminant ability of the learning models. It also maximizes the between-class feature matrices S_Bs of transformed features and diagonalizes the individual S_B to ensure well class-separation. Finally, it finds two sets of feature transformations for each pair input by simplifying Eq.(4.8a)-(4.8b) as-

$$\bar{u} = A_x X; \bar{v} = B_y Y; (W_i = A_x, B_y; X_1 = X, X_2 = Y) \tag{4.12}$$

The transformed diagonalized features have high correlation only with their own features and are subjected to feature fusion strategy similar to the CCA. Similarly mmDCA evaluates other sets of feature using DWT-dMV features and finally, fuses using summation

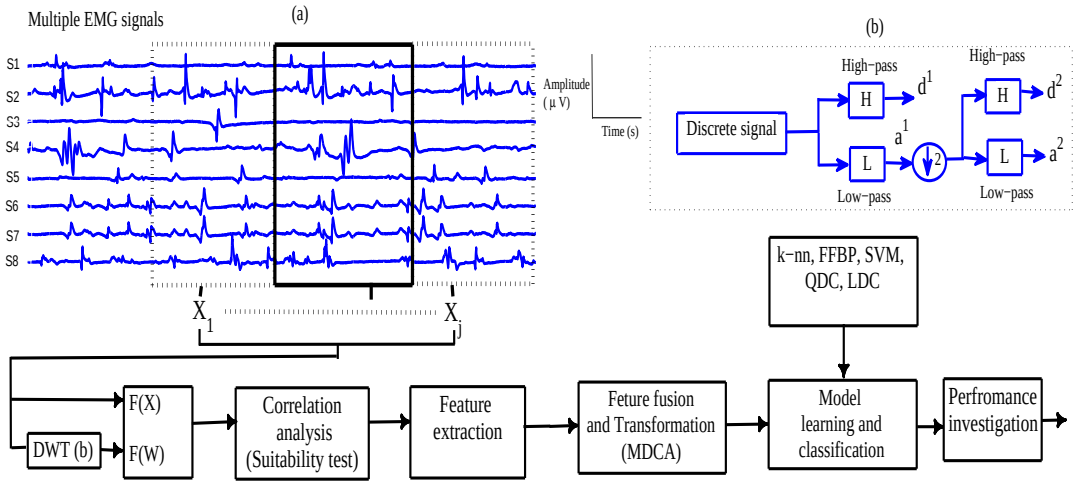

Fig. 4-3: (a) EMG signals S1-S6 of a given process are arranged in order to find the features $X_1, \ldots, X_j$ for mmDCA analysis. Features are then transformed using DWT to obtain $X_1(w), \ldots, X_j(w)$. F(X) and F(W) represent two sets of independent features which are subjected for the model in the second layer, and (b) two-level decomposition tree of the DWT is used to formulate F(W).

technique as shown in Eq.(3.14a) to find single feature vector as-

$$Z_{ij} = \sum_{t=1}^{4} T_t(\bar{u}, \bar{v}) = A_x^T X + B_y^T Y + C_{u1}^T U_1 + D_{u2}^T U_2 \tag{4.13}$$

Terms $T_t(.)$ and Z_{ij} indicate canonical variate and generalized canonical correlation discriminant features (gCCDF) of $\{i, j\}th$ dMV pair; C and D are weight vectors of wavelet domain dMVs. This synchronized feature mode improves the learning ability of models [7]. The discriminant features are evaluated for all consecutive input pairs of dMV. Compared with DCA, the proposed mmDCA copes multi-domain large-scale inputs into gCCDFs, based on which more favorable statistics can be evaluated to effectively predict and diagnose various nonlinear processes.

In evaluating DWT features, iterative decomposition is restricted upto second level as per guidance of possible frequency range of signals and further, higher frequency d^2 components are discarded from the analysis. Fig. 4-3 shows the two-stage feature evaluation scheme and DWT components for each study group C is shown herein. It is further aim to establish correlation of MV or dMV features for validation of proposed feature extraction scheme and to extract suitable features for analysis.

4.5 Results and discussion

4.5.1 sMV and dMV feature evaluation

In order to find the features according to our proposed model as shown Fig. 4-1, the study subject groups are divided into a number of subgroups in Table 4.1. Then, sMV

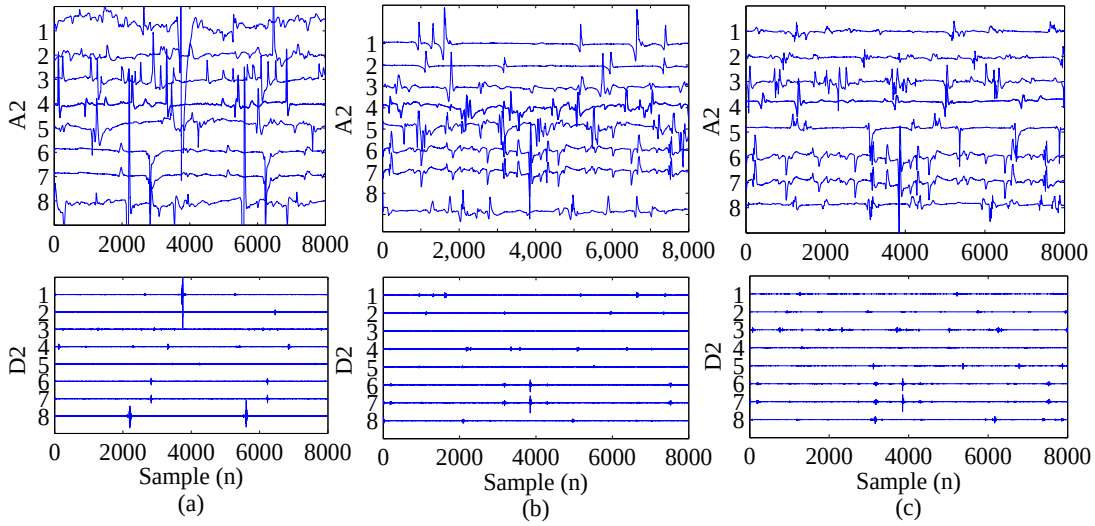

Fig. 4-4: Second order wavelet components (A2, D2) for three categories subgroups of ALS, a) subgroup 1 (age: 35-52), b) subgroup 2 (age: 53-61) and c) subgroup 3 (age: 62-67)

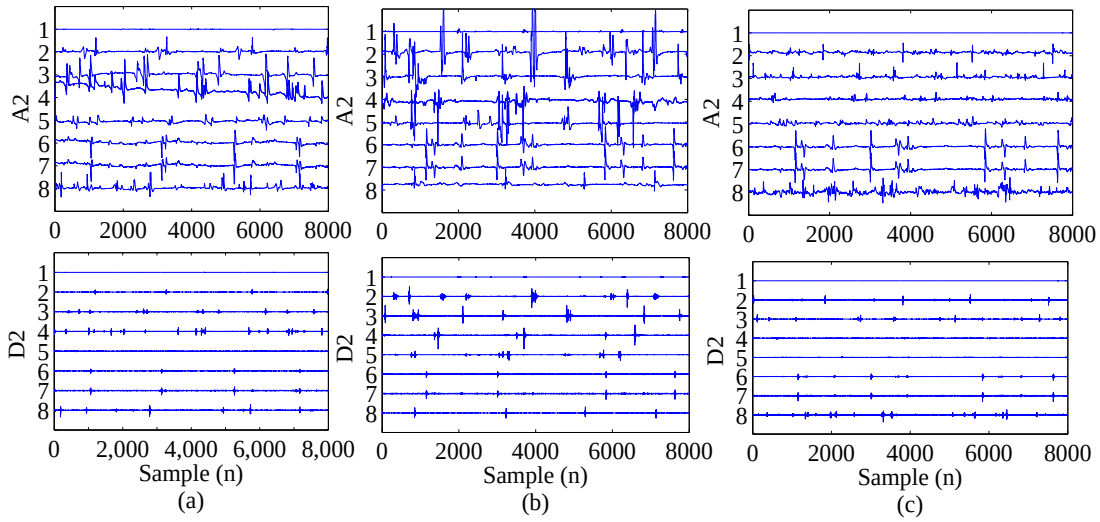

Fig. 4-5: Second order wavelet components (A2, D2) for three categories subgroups of myopathy, a) subgroup 1 (age: 19-26), b) subgroup 2 (age: 28-41) and c) subgroup 3 (age: 44-63)

and dMV features are evaluated for analysis and fusion to find the discriminant features.

Each subject group is divided into three subgroups (i.e., m=3) according to the age of subject as outlined in simplified Table 4.2. Each subgroup comprises of two subjects with large signals. Three sMVs corresponding to three age-based subgroups are evaluated. From each sMV, eight dMVs (i.e., j=8) are obtained as shown in Fig. 4-2. This way, 24 dMVs are evaluated for each study group. In the same way, an equal number of dMV are evaluated using DWT-sMV. Fig. 4-8-Fig. 4-6 show various DWT components of various subgroup signals. In finding the sMV, p=8 signals are selected taking two signals from each subgroup and remaining two signals are randomly selected from the subgroups ($= 2 \times 3 + 2$). This optimal choice is estimated using statistical test ($p < 0.05$) during performance assessment. The dimensions of sMV and dMV are 8×258000 and 8×32250 respectively.

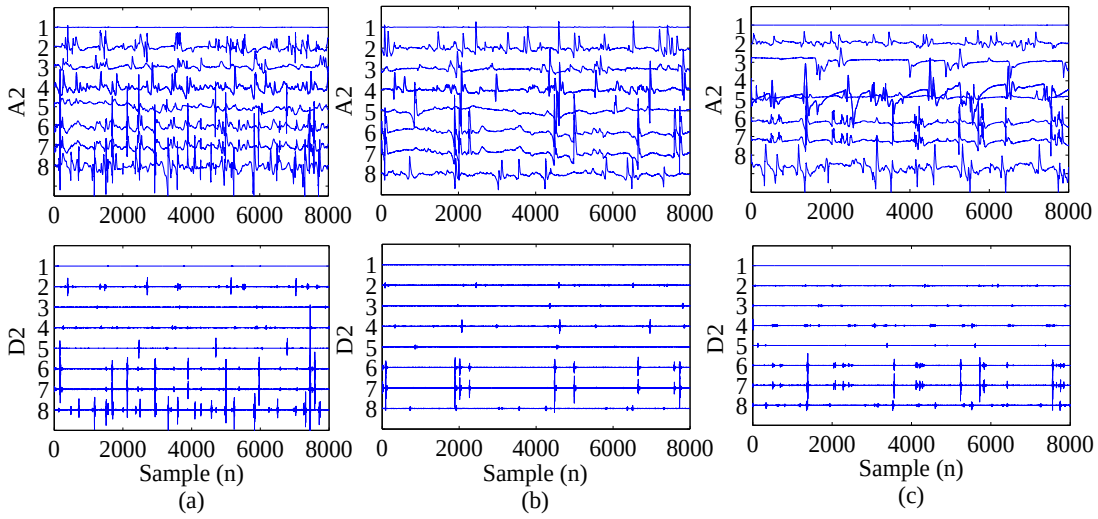

Fig. 4-6: Second order wavelet components (A2, D2) for three categories subgroups of normal, a) subgroup 1 (age: 21-26), b) subgroup 2 (age: 27-29) and c) subgroup 3 (age: 29-37)

Table 4.2: Evaluation of sMV and dMV for three processes

C/Gc	Age range [Year]	sMV	dMV
ALS	35-52, 53-61, 62-67	$X^{1(1)}, X^{2(1)}, X^{3(1)}$ (3)	$3\times8 = 24$
Myopathy	19-26, 28-41, 44-63	$X^{1(2)}, X^{2(2)}, X^{3(2)}$ (3)	$3\times8 = 24$
Normal	21-26, 27-29, 29-37	$X^{1(3)}, X^{2(3)}, X^{3(3)}$ (3)	$3\times8 = 24$
C/Gc= 3	Total m= 3+3+3	Total=3+3+3	Total = 72

4.5.2 Suitability of features: correlation analysis

In correlation based process measurement method it is necessary to ensure whether selected features are correlated or not [6] to get an idea about the suitability for such measurement. Therefore, we estimate the correlations among various dMV features for three respective subgroups. Correlation measures are being carried out for age-based sMV features as shown Fig.4-7(a). It indicates that age-based sMV features have less correlation due to wide dissimilarities of the data pattern in features. In other words, it infers that observations forming the sMVs highly depend on subject age and profile of disorder. Fig.4-7(b) indicates that dMV features of respective groups are highly correlated and of similar nature except for the normal case. Thus it remains in favor of the feature generation scheme of the dMV advocated in this work. In estimating the correlations eight observations are considered. Two from each subgroup and other two are randomly selected from the subgroups for this analysis ($= 2 \times 3 + 2$). This optimal choice is further supported by statistical test ($p < 0.05$) during performance assessment.

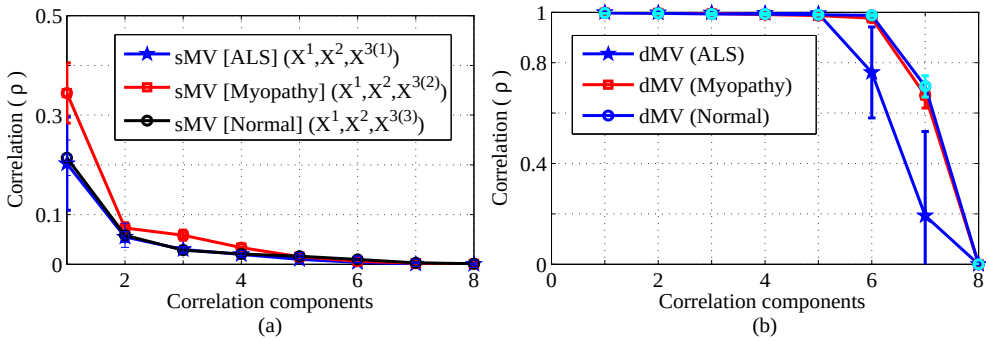

Fig. 4-7: a) Mean correlation among sMV feature within same subject groups. The sMV features, $X^{m(1)}$, $X^{m(2)}$, and $X^{m(3)}$ fall in the age range of [35-52, 53-61, 62-67] for ALS, [19-26, 28-41, 44-63] for myopathy and [21-26, 27-29, 29-37] for normal, and b) Mean correlations among the consecutive dMV features for three classes.

Table 4.3: SM evaluation in terms of l_1 and l_2. Highest values are indicated by boldface.

Methods	l_1	l_2
Conventional PSD based methods [136]	0.69	1.82
STFT and RVM [136]	0.74	4.57
Q-factor wavelet and spectral features [137]	2.0739	148.4636
Our previous model	**3.987**	**149.0125**
mmDCA	**8.675**	**155.321**

4.5.3 Order selection and separability evaluation

Using Eq.(4.13) the features are evaluated for all values of m. Four gCCDFs are evaluated for each value of m and consequently twelves gCCDF are evaluated for each categorical input process. Thus, total gCCDFs is 24, twelves from each domain analysis. In other words, domain $\times m\times$ pairs= $2 \times 3 \times 4$=24. However, the discriminant features are evaluated using correlation threshold at six based on the analysis as in Fig.4-7. Then, features are transformed using the linear transformation Eq.(3.20). The group-specific mean gCCDFs are further subjected to linear discriminant analysis that minimizes within-class variance and maximizes between-class variance to attain optimal features [16]. A ANOVA test is carried out and shows that all the selected features have p-value less than 0.05. Most of the selected have p-value $< 1 \times 10^{-4}$ and other have very close to zero. The p-values of features obtained without aforementioned linear transformation are also statistically significant.

Fig.4-8 shows the box-plot feature distribution that have discriminant capability. Further, SM in terms of l_1 and l_2 is also calculated in Table.4.3 using Eq.(3.27a)-(3.27b) (Section 3.5.1.3). As is evident, the SM is higher than previously described approach in Chapter 3 and in spite of being overlapping, feature vectors provide good inter-class variation and intra-class separability. Thus, the mmDCA provides better discriminant features to achieve robust performance while integrates with classification models.

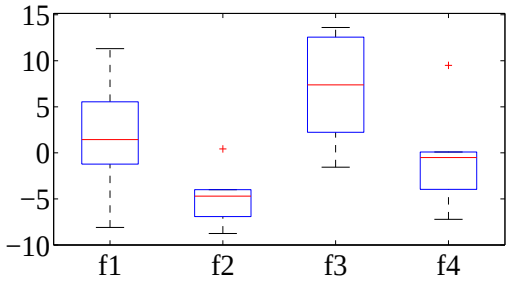

Fig. 4-8: Box-plot feature distribution of selected features for classification task using mmDCA.

4.5.4 Performance of mmDCA on real-time datasets

For performance evaluation, the dataset EMG_{N2001} is partitioned into three subsets-training set (50%), validation set (25%) and test set (25%). The performance is investigated by integrating simple classifier k-nn with the best feature combination obtained using inter-cross validation technique. The classifier has been supplied with discriminant features to be assigned any one of three input processes or patterns. With this combination, the mean accuracy obtained is 98% with a specificity of 99.30% and sensitivities of 97.61% and 97.14% (SnA and SnM). The results showed a fairly good linear separability of data or features obtained from the proposed mmDCA scheme. Furthermore, to show capability features in large-scale information, two additional simple prediction models-normal densities based linear classifier (LDC) and quadratic classifier (QDC) are also integrated with the scheme. Table.4.4 shows the performances of various prediction models in terms of markers with manual annotations. It shows improved prediction performance with higher values of other parameters on the same data sets. Furthermore, results are very close to each other which evince the superiority of the mmDCA. The average run times shown in Table.4.4 are also reasonable.

In multiple process prediction systems, the higher value of the individual parameter is essential in addition to the accuracy. For example in our case, specificity is used to show prediction rate of the negative case (normal process) while sensitivity is used to show prediction rate of positive cases (ALS and myopathy). However, for all positive cases, all these parameters give the details about the model performance. The proposed method performs very well irrespective classifies, which are presumably due to the fact that they significantly remove the redundancy in feature data set and establishment of high quality low order features that have low variance in feature registration stage. The inclusion of class association of features, the feature space have inter-class marginal in terms of SM results in improvement of accuracy as compared to our previous method. The overall outcomes are uniform and reasonable for practical implications. It is, thus, conclude that use of multi-domain-multi-view fusion models are promising and significantly improves the reliability.

Table 4.4: Mean performances of various integrated data-driven schemes and processing time.

Models	Sp_μ (%)	SnA_μ (%)	SnM_μ (%)	AC_μ (%)	Run-time (in seconds)
mmDCA+kNN	99.30	97.61	97.14	98.00	45
mmDCA+LDC	99.43	98.09	98.57	98.60	23
mmDCA+QDC	99.58	98.50	97.59	99.03	33
mmDCA+QDC [EMG_{GNRC}]	100	100	100	100	21

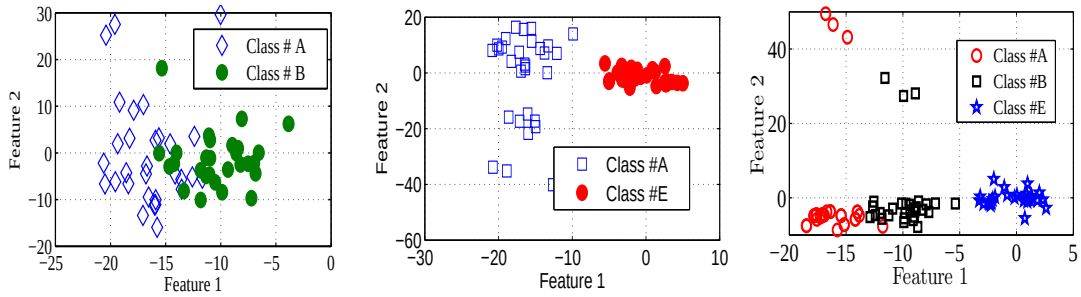

Fig. 4-9: Scatter distribution of features of the mmDCA for various groups features, A, B and E (#*A*-ALS, #*B*-Myopathy and #*E*-Normal).

4.5.5 Reliability and Scalability of the mmDCA

The features distribution in Fig.4-8 and Fig.4-9 obtained from the mmDCA and subsequent performance measurement in Table.4.4 clearly evince the reliability of the mmDCA based assessment scheme. As is evident, the scatter distributions of various combination of features have higher separation margin (i.e., $l_1 = 8.675$ and $l_2 = 155.321$) and low feature dispersion which are fundamental requirements for accurate process prediction [48]. Large coverage of available input space and extracted low order feature structure carrying more information about the process with significant p-statistic make the mmDCA more reliable for real-time assessment and accurate prediction. It is to be mentioned that the proposed model has well-defined feature extraction and fusion strategy that significantly improves the learning ability of model. Furthermore, free from assumptions and ease of assessability promote to implement for many industrial process assessment schemes such as-quality assessment and prediction [147, 148], false prediction [149, 150], diagnosis [151, 152] in various domains.

The scalability of the mmDCA is evaluated dealing with a new dataset of EMG_{GNRC} that is not used for aforementioned measurement. The obtained features using EMG_{N2001} is used to evaluate the performance over the EMG_{GNRC}. For each input process, a set of gCCDFs are estimated using method in section 4.3 taking three sets of randomly selected samples from EMG_{N2001} and their means are used for analysis. The accuracy obtained by the optimal model with this feature space is 100% with a specificity of 100% and sensitivities of 100%. The optimal model performance is due to low order discriminant features and presumably due to small database of EMG_{GNRC}. It

is worth to be mentioned that despite reduction of noises up to a significant level, some inherent noises are obviously in signals of EMG_{GNRC}. However, these noises have low impact on model performance as per the principles of CCA and mmDCA. To further ascertain the effectiveness of the proposed methodology, measurement is also carried out with two data sets obtained from EMG_{GNRC}. However, in both cases similar results are obtained indicating the robustness of the mmDCA.

4.5.6 Comparative analysis

This section briefly outlines various EMG diagnosis methods and provides the comparison of results with our results. It further focuses on addressing various nonlinear process assessment methods widely used in industrial domain to highlight the efficacy of the proposed method.

4.5.6.1 Diagnosis of neuromuscular disorders

Inherent limitations, theoretical bottlenecks and computational complexities of various methods have already being addressed in the Section 3.5.2.3. For instance, the method in [76] requires optimal radial basis kernel function parameters which are tricky. Some method takes large memory and slow down the process [80] while others, specifically DWT-method and MUAP-based methods often face difficulties in finding proper wavelet coefficients that accurately match dominated MUAP. Some model provides good accuracy level, however in some case it fails to maintain balance values of other markers, e.g., method [154] reported Sn of 64.29% for neuopathy with high accuracy. Despite potential impacts, model inherent limitations make them unsuitable for real-time applications. Besides, they did not address the use of large data using feature fusion model. However, big data usage is essential for reliable healthcare intelligence [2]. Most of these studies focused on statistical models to obtain key statistics for process monitoring. Use of large data using fusion model is a good choice for reliable system.

In the context of performance markers, our method shows promising results as compared to many state-of-the-art method. Furthermore, it is simple and easy to understand for different real-world applications which are similar to our studies. Effective use of data through use of discriminant formulation leads to a promising inference system that outperforms many reported methods outlined in Table.4.5. Thus, use of nonlinear data associated with medical process accomplished in the experiment and logical inferences also ensures the possible real-time implementation of our algorithm for various complicated process monitoring in industrial applications. Such algorithm has potential values in the process industry and medical field.

Table 4.5: Performance comparision with various EMG support models in terms of biomarkers.

Methods	Year	Class	Sn1 (Sn2) (%) (%)	Sp (%)	Ac (%)
Subasi [70]	2012	3	94.50, 95.75	95.25	97.67
Subasi [155]	2013	2	98.00, 99.00	95.25	97.40
Kamali *et al.* [57]	2014	3	95.00, 96.00	97.60	97.00
Gokgoz *et al.* [76]	2014	3	95.30, 95.67	96.33	92.55
Subasi [79]	2015	3	98.25, 99.00	93.75	97.00
Gokgoz *et al.* [80]	2015	3	99.58, 95.66	94.75	96.67
Naik *et al.* [23]	2016	3	94.00, 96.00	100.0	98.00
Hazarika *et al.*	2016	3	98.00, 98.00	99.00	98.80
Proposed [EMG$_{N2001}$]	–	3	**98.50, 97.59**	**99.58**	**99.03**
Proposed [EMG$_{GNRC}$]	–	3	**100, 100**	**100**	**100**

4.5.6.2 Industrial process management systems (IPMS)

In order to highlight the applicability of the proposed method in various industrial process assessments, this section outlined many such methods in Table.4.6. It includes locally weighted total projection to latent structures [149], advanced partial least squares [151], intelligent particle filter [150], key performance indicators (KPI) [156], modified partial least squares [147], fuzzy positivistic C-Means [157], parameter estimation [152], modified orthogonal projections to latent structures [148], probabilistic-PCA [158], multivariate statistical process monitoring [159], kernel PCA [160] and kernel PLS [161] for industrial process assessments. Many other similar methods in [162–164] are also remarkable. Data-driven models that reduce the burden of analytical model rely on the measurable data [165]. However, the applicability of various models is limited by their inherent pitfalls. The linearity assumption in PCA and PLS limits the nonlinear applications [165]. The methods [160] and [161] could improve the performance in nonlinear process prediction, however, selection of kernel function and parameter determination are difficult. Some models employed KPIs using PLS [166] and locally weighted projection [165] from the available measurement for abnormality prediction. These methods popular in literature for application specific task or domain specific task, however, they did not address the fusion scheme to reduce the feature dimensionality. Instead they focus on statistical makers based on diagnosis can be made. In that context, the proposed mmDCA is more reliable and appropriate choice.

4.6 Performance variations

This section highlights the variations in outcomes obtained from various data-driven based learning schemes with used of proposed large-scale information embedding strategies in terms of comprehensive markers - accuracy, specificity and sensitivities. Chapter 3 provides two independent formulations with use of generalized CCA while this chapter

Table 4.6: State-of-the-art industrial process assessment and dignosis algorithms.

Methods	Assign tasks	Year
Shen *et al.* [149]	2017	Abnormality detection
Xiaochen *et al.* [151]	2016	Prediction and fault diagnosis
Yin *et al.* [150]	2015	Fault Detection
Shardt *et al.* [156]	2015	Fault detection
Wang *et al.* [147]	2015	Quality-related fault detection
Yin *et al.* [157]	2015	Fault detection and isolation
Zhai *et al.* [152]	2015	fault diagnosis
Yin *et al.* [148]	2015	Quality-related fault detection
Zhu *et al.* [158]	2014	Process monitoring (outliers/missing data)
Haghani *et al.* [159]	2014	fault detection

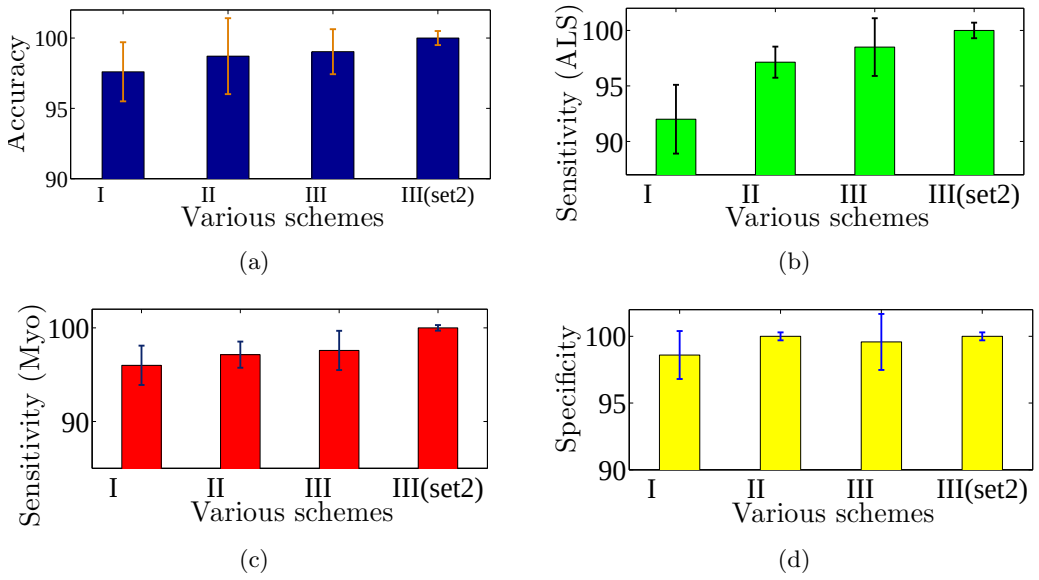

Fig. 4-10: Variations in terms of performance markers, a) accuracies of proposed feature fusion schemes I, II and III, III(set 2) b) sensitivity of ALS, c) sensitivity of myopathy and d) specificity of negative case. Here I, II and III stand for three schemes - CCA, multi-view CCA and multi-view multi-domain DCA or DCA-PAIS respectively.

provides generalized DCA scheme. Each scheme assessment comprises of a set of these markers. Further, the optimal model performance is also assessed with new real-time data set EMG_{GNRC} (set 2) so as to ensure the reliability and scalability.

Fig. 4-10 shows the results obtained from three schemes. Variation in mean accuracies of the schemes with three feature fusion strategies I, II and III, is shown in Fig. 4.11(a). Fusion schemes I and II are based on generalized CCA while III is based on generalized DCA. The scheme III (set 2) represents the reliability and scalability of the optimal scheme III. It is seen that accuracies are very close to the optimal value and variations are minimum among the various schemes. Thus, it infers that the proposed feature fusion based schemes suitable for covering large-scale information from available inputs. CCA generalized scheme II improves the accuracy in comparison to the I. However, DCA scheme III outperforms over I and II. It is thus clear that CCA and DCA

based learning strategies are the suitable choices for multi-task classification problems irrespective of their small variations in accuracy. It is presumably due to minimum feature biasing in the frameworks. In this context, the reliability and scalability analysis shows no variations in the respective analysis as seen from Fig.4-10. Sensitivity variations are relative to the overall classification accuracy of respective models. However fluctuation in sensitivity measuring the predicting capability of positive case ALS is higher than that of myopathy (Fig. 4.11(b)). In case of myopathy this variation is minimal among three schemes as in Fig. 4.11(c). Similar to above discussion the specificity variations can be explained.

The performance of data-driven techniques relies on the initial framework that captures the information of underlying processes. Furthermore, it also depends on the degree of freedom of parameters inherent in the models. Feature fusion based framework is the most intuitive in the sense that it uses multiple observations to extract usable information in terms of feature sets which contains much richer information about the underlying process. This undertaking work therefore mainly focused on information management system with special emphasis on feature extraction and fusion strategy and integrated simple prediction schemes for the classification task. Due to dimensionality concern, this study integrated simple classification models to cope obtained feature space from the feature fusion frameworks. Such analysis further ascertains that whether the proposed feature extraction strategies are capable of covering wide variety large-scale information in real-world applications. The significant performance of specific algorithm may not ensure the efficacy of feature space. Because of the small dimensionality of obtained feature space, it is easy to adopt various learning schemes for prediction of the various non-linear medical process. The capability of various feature concatenation frameworks through the assessment of individual measure is explored in aforementioned discussion. On the basis of outcomes the optimal feature fusion framework III, i.e., multi-view DCA or DCA-PAIS, has been selected and is subjected to analyze the outcomes variations with changes of classifying models. As a consequence three simple classification models - k-NN, LDC and QDC were integrated with the III, referred to as A, B and C and the respective outcomes with the real-time data set EMG_{N2001} are presented in Fig. 4-11 in terms of mean markers. In this context, the scalability of the III was investigated dealing with new data set of EMG_{GNRC} that was not used for aforementioned measurement and outcomes are shown herein (C(EMG(II)). It further ensures the effectiveness of the III. It is seen that variations in outcomes in terms of markers are quite insignificant. In another word they are very close to each other which evince the superiority of the III. Also, comparison of the average run times is reasonable in the context of this analysis. From this analysis, it is clear that classification framework highly depends on an efficient strategy to incur large information from available inputs in terms of low order statistics rather than classifier. However, integration of simple prediction scheme is more suitable for such assessments.

The proposed data-driven model incorporated feature fusion frameworks gen-

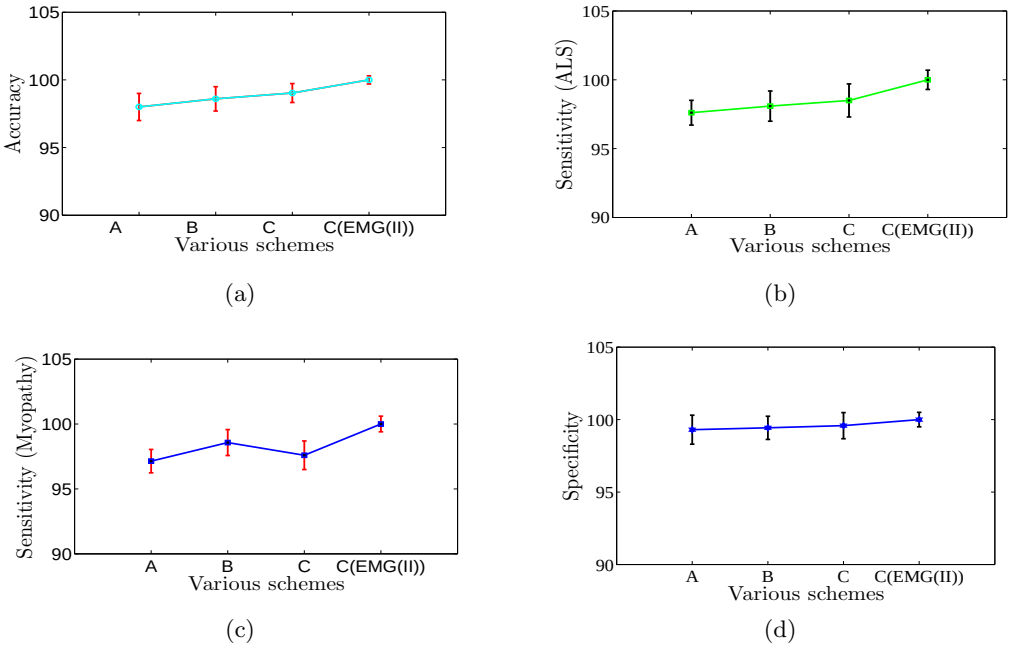

Fig. 4-11: Variations in performance with optimal feature fusion scheme III as obtained from above analysis- a) accuracies of learning schemes - A, B and C b) sensitivity of ALS, c) sensitivity of myopathy and d) specificity of negative case.

eralizing CCA and DCA for nonlinear process is investigated. The optimal results and variational analysis among various feature fusion based schemes shows that the DCA based multiview scheme, also known as the DCA-PAIS scheme is capable of extracting information from the available measurement and provides comprehensive statistics that accurately diagnose the processes. In the context of achieving promising results and reliability and scalability, the scheme III is more suitable in comparison to other models.

4.7 Conclusion

This chapter addressed a more generalized multi-view feature fusion data-driven model based learning scheme for EMG diagnosis. A new feature extraction strategy and an efficient multi-domain feature fusion-based mmDCA generalizing discriminant correlation analysis are proposed and investigated with EMG data sets. This scheme is capable of extracting information from the available measurement and provides comprehensive statistics that accurately diagnose various non-linear medical abnormality patterns. The core focus of the method is multi-view feature fusion framework to cope large information in terms of low order statistics for the complete understanding of underlying phenomena occurring in real-world. Successful implementation of the algorithm to diagnose neuromuscular disorders, results and discussion and comparison with the state-of-the-art methods reveal the effectiveness of the mmDCA. It significantly improves the feature biasing issue which is common aliment in multi-task learning. In assessing the per-

formance, other important issues like dimensionality, reliability, easy accessibility and robustness for viable implementation are addressed. Further, statistical tests-one way and two-way-ANOVA are also performed. Thus, the proposed approach is efficient in managing large-scale data and helps alleviating the computational and theoretical bottleneck, and can be applied to nonlinear process monitoring. The fully automated system with such efficient algorithm implementation aids physicians, large scale diagnosis research as well as portable device implementation for detecting alarming trends in the health status of users.

To the end, the chapter also highlighted a few important issues. Albeit significant of performance, many data-driven models fail maintaining the consistency with wide variety of data sets. Involvement of multiple stages for feature extraction can be potential source of error. It is seen that many prior methods used low order statistical measures such as μ, σ in case of high dimensional features, for effective class distribution. In case of close input data distribution of various classes such parameters may not provide meaningful measures. In that case our approach is an attractive choice for processing and analysis of biomedical signals. It controls the complexity, over-fitting of models efficiently. These factors in conjunction with aforementioned advantages have pushed toward attaining higher accuracy of mmDCA based learning.