Chapter 5

Conclusion and future direction of research

5.1 Conclusion

The intuitive nature of advanced feature fusion techniques in real-world scenario motivated our approach to incorporate multiple views information of input sources with an aid of statistical models for non-linear process monitoring and diagnosis. In this thesis, we dealt with such methodology integrating different form of data in order to provide an automated reliable framework for real-time implementation. As argued in this thesis, the methods uniquely provide the link to account the abundance of diverse information from various forms of input measurements or data for the complete understanding of undertaking phenomena. The potential impact of the gains through the assessment of performance in analyzing nonlinear medical process is also high for many industrial process monitoring and diagnosis, environmental, psychological problems that cannot be solved based on single-set measurement. This thesis covers a number of topics in the context of deriving feature fusion based learning by generalizing the CCA and DCA. The discussion of these topics was distinguished in various aspects: multi-view feature creation, fusion strategy, theory, algorithm, real-time implementation with benchmark data, comparison analysis with existing methodologies.

In the initial stage, many state-of-the-art EMG classifications and feature fusion based algorithms used machine learning applications are reviewed. Further, the limitations and the possible aspect of feature fusion based algorithms in the real scenario are also explored. The prospective of feature fusion based algorithm enabled us to plug the CCA and DCA in supervised learning problems to develop new algorithms incorporating large scale input information space. To tackle such problem and to avoid the computational burden, a set of strategies were proposed and investigated with aid of optimization techniques to simplify the problem. One of the main findings was that

multi-view formulation frameworks are very efficient in the context of large information embedding and in avoiding feature biasing in the initial framework of learning model.

5.2 Achievements

The aim of this thesis is to achieve deep understanding of the nature of variations inherent in bio-medical signals and insights about the underlying phenomena associated with the signals. Furthermore, information extraction relying variations of signals and development of an appropriate feature interaction strategy in the initial framework of data-driven models and sound feasible algorithm that can overcome the theoretical and computational bottleneck of many earlier methods in medical domain are another major goal of this thesis. The thesis argued various aspects of theoretical formulations and results obtained in various sections. A list of main scientific contributions of this projects is briefly outlined in the following.

5.2.1 Specifying high dimensional multi-view feature

Three independent multi-domain multi-view feature generation schemes were addressed with special emphasis on feature biasing issue related to the initial framework of learning since the model performance highly depends on appropriate utilization of inputs to extract suitable feature sets rather than the choice of classifier [92]. Various strategies in order to create multi-view input features have been provided and feature extraction schemes have been modeled as logical predicates using CCA, mCCA. At an initial phase, the variability and stability of within and inter-subjective signals to support the adopted model and to select appropriate features have been investigated and comprehensive features are evaluated. Such large-scale information extraction models have been shown promising to avoid feature biasing issue. Further, dimensionality reduction is explored. The choice of such features is due to fact that the multi-view feature derived from multiple signals should have most physiologically relevant information and energy contents. It thus contains much larger information compared to the specific signal feature.

In another approach, a sub-multi-view feature based scheme has been modeled. Unlike other approaches, this scheme focuses on use of generalized DCA that has a similar theoretical formulation with the CCA with exception of FDF. Improvement in generalized ability of the weighted feature space has been examined. The reliability of this features in terms of quality information and compactness with reference to previously evaluated features are examined. The complexity of this scheme is examined in the context of the aforementioned strategies which reveals the unbiased property of this scheme. Such models are efficient in managing large-scale data and representing underlying process in terms of low order feature subsets for classification task.

5.2.2 Feature fusion model for discriminant representations

Using CCA based feature fusion models-parallel and serial fusion two algorithms for discriminant representation of transformed features that corresponds to various input biomedical processes, have been introduced. Furthermore, the global feature descriptors in terms of mean CCDF and gCCDF have been provided. However, the dimensionality of features obtained using two fusion models is different. So, to avoid this issue, the low order fusion strategy is used throughout the analysis. Two data-driven approaches-mCCA and MVDCA address the problem of high dimensional input space and well-facilitated the multi-view information or knowledge embedding for effective performance model evaluation. Furthermore, the model-related issue like complexity, feasibility, reliability, repeatability etc. of various models have been investigated.

5.2.3 Evaluation of model performance

In assessing the performance, various classification models are integrated with feature fusion based data-driven models-(i) k-NN is integrated with the CCA, (ii) k-NN, QDC, LDC with the mCCA under two strategies S-I and S-II, and (iii) k-NN, QDC, LDC with the MVDCA. The performance and efficacy of data-driven models are explored in the context of results obtained in terms of comprehensive performance markers-accuracy, specificity, and sensitivity on various combination of two real-time databases-EMG_{N2001} and EMG_{GNRC} which consist of three category subjects-ALS, myopathy and healthy control subjects. Improvement of the performance of various generalized version is also examined. Furthermore, to ensure the suitability of the selected feature space, the variations in output measures in terms of statistical parameter-standard deviation (σ) are investigated which shows low variation in outcomes. Thus, such large-volume information embedding models are promising for biomedical signal classification. The outcomes are briefly outlined in the following.

Case (i): the performance is evaluated using three type of features-Type-I, II, and II over two-class and three-class subject group data. In two-class data group, the accuracy obtained using Type-III was 98.10% with specificity and sensitivity of 99.65% (ALS-Normal) each and in three-class groups, the accuracy, sensitivities and specificity - 97.3%, 92%, 96.0% and 8.60% with Type-II over EMG_{N2001}. However, it achieved the optimal level of parameter values (i.e.,100%) over EMG_{GNRC} using Type-II.

Case (ii): Classification performance is investigated over the EMG_{N2001}, EMG_{N2001}+EMG_{GNRC} and EMG_{GNRC} independently with S-I and S-II. The overall mean accuracy obtained using S-I is found to be 96.8% with a specificity of 98.8% and sensitivities of 96.6% and 97.6%. With S-II the system shows the highest performance over the database EMG_{GNRC}. However, the mean accuracy is 99.4% and the sensitivities and specificity are 99.5%, 97.7% and 100% respectively. In all the cases the classifier

achieves promising results with S-II in comparison to the S-I. The reliability of the MV based measurement scheme is demonstrated in the context of the improved results and specifically low variance in various outcomes obtaining from different strategies.

Case (iii): The scheme MVDCA based scheme with QDC achieved mean accuracy of 99.3% with a specificity of 99.58% and sensitivities of 98.50% and 97.59% over EMG_{N2001}. However, the same approach achieved optimal parameters' value over EMG_{GNRC}. Furthermore, various models were in conjunction with the MVDCA used to demonstrate the variations in outputs.

5.2.4 Reliability, scalability and variability analysis

The reliability of the advocated multi-view based learning schemes has been demonstrated in the context of the results obtained using various classification models. The extended version of the initial model as stated in case (i) is subjected to this measures. In case (ii) DA with linear and quadratic achieve mean accuracy of 97.2% with a specificity of 99.3% and sensitivities of 97.5% and 96.4% with overall variation in terms of σ was of 1.3% (accuracy). In case (iii) these values are 98.3% with a specificity of 99.4% and sensitivities of 97.6% and 97.7% with overall variation in terms of σ is of 0.5% (accuracy) with reference to the best value in this scheme. Thus, the reliability of various schemes is demonstrated. Furthermore, the scalability is examined using features from the training of model with one population are applied to classify new dataset. The features extracted using the datasets of EMG_{N2001} is used to estimate the performance over EMG_{GNRC}. Such analysis evidences that proposed algorithm is robust and reliable for diagnosis of neuromuscular disorders.

The results obtained using our methods compared with many relevant state-of-the-art methods for better explore the efficacy. The topic narrated in this thesis mostly comes from the area of pattern recognition in machine learning. In this field, things are changes rapidly and new algorithm always look better than old models. The most promising method shown is one which integrates statistical data-driven with classification model. Despite massive advancement in this field, challenges are there in such models. The true merit of the proposed algorithm lies in the fact that this algorithm provides accurate predictions of various complicated medical processes in the burning field, which lead us to the ultimate goal of computational intelligence. To the end, it is specifically helpful for clinical applications and device implementations. The proposed operational framework is not an alternative approach to replace the doctor in any way but can support and enhance the diagnosis process so as to take early alarm of patient health status for quick diagnosis.

5.3 Future direction of research

The research presented in this thesis opens the doors to many interesting perspectives in several directions.

Data or feature fusion model enable us to cope multiple modalities heterogeneous measurements or sensors outputs for logical inferences. Apart from biomedical applications already mentioned in this thesis, it also has a wide range of applications in many engineering fields. For example multi-modality analysis such as EEG and MRI, speech and acoustic signal processing, multi-sensor measurement-based analysis, robot control which typically use various sensor modalities-vision, audio, range sensors, etc. to capture information. Various sensors produce different kind of output or information and ensemble of these sensors responses is essential for real-world applications. However how to utilize these sensors information effectively with high reduction of the negative parameter becomes a critical issue for the deployment of mobile robots. Such measurement is not limited, many medical diagnosis strategies deploy multiple modalities measures for final interpretation of disease. As a consequence in recent time, data fusion or feature fusion algorithms have been considered as hot research topic and played a principal role in the acquisition of more accurate and reliable information for the last two decades. Many of fusion models achieve remarkable success in many real-world applications including multimodal fusion [1], multisensor data fusion [1], structured data fusion, audiovisual fusion [7], image fusion [28] and multiview data [48]. However, multiview feature fusion based assessment scheme presented in this thesis is a new concept and have not been used in aforementioned methods. Therefore, the methods with various strategies introduced in this work presented may be applicable to develop a new paradigm of industrial process control, fault-identification, medical diagnosis and large-scale research.

The thesis presented an application of combined attributes in terms of the feature from true multiple data interaction model. The improvement in performances as indicated by the results and comparative analysis with domain-specific models could inspire us to apply such model for analogues applications such EEG, ECG and multimodal application such as fMRI, sMRI, and EEG. The algorithms CCA, mCCA and DCA based learning proposed in this thesis are technically sound feasible for clustering applications. In literature, there are very few relevant methods which used multi-view feature fusion models. Furthermore, any CCA generalized versions such as MICCA, MCCA LDCCA, and kernel version of LDCCA were seen which may be suitable for such applications. How to incorporate such models in supervised learning applications is still an ongoing problem.