Chapter 1

Introduction, motivation and challenges, and overview of state-of-the-art methods

1.1 Introduction

Information fusion, specifically feature fusion embeds vast amounts of information or data from single or multiple modalities profiles to provide an efficient learning and reliable inference that can greatly impact the application's success. Furthermore, in order to meet ambitious performance requirements, use of feature fusion technique in machine learning applications is essential, since it enables true interaction among multiple sets of data or signals and provides low order features. These features have much discriminant information about the underlying phenomena. As a consequence, there has been a growing interest in feature fusion in machine learning, which is a contemporary topic rooted in engineering. The algorithm with such intelligent signal processing technique enhances the reliability and enables embedding in decision making platform such as microcontroller (μC) for portable device-based online process monitoring and diagnosis.

Building a multi-view or multiple modalities information fusion based learning that works well across multiple fields is an arduous task and specifically challenging in the medical domain. Term multi-view here indicates different forms (i.e., input features) obtained from single or multiple modalities profile. Real-world measurements often involve a large number of diverse sensor signals, each of which carries information of different degree. Therefore, to develop an efficient learning algorithm, it must cope with integrating information from large-volume data such as multi-view in an intelligent way. Besides mandatory requirements of multiple modalities profiles, most of the early work often manage to provide a simple model with some specific information for specific applications that works well, however, work poorly across multi-domain applications.

Nonetheless, all of them encourage the researchers to carry out new work or modification on the existing methodologies so as to find the optimal solutions and implementation in real-world scenario. Importantly, simplified models are well adapted due to ease of understanding and accessibility, wherein inconsistency and non-reliability are two inherent pitfalls. Therefore, trading off these markers is usually not reasonable in many domains, albeit its remarkable reinforcement towards knowledge community.

In order to explain the problem, we consider two sets of electromyogram (EMG) signals collected from two subjects under same disease group with proper supervision of expert personals. It is the fact that two sets are significantly different due to inherent non-stationary nature with adaption to noise and thereby, neurophysiologists assess multiple sets with their knowledge which is achieved at the later level of service. Furthermore, in the early stage of disease, changes in recorded signals are not prominent. In some cases they fail and have to wait for complete development of symptoms to make a correct decision. Such belated diagnosis process increases the cost burden and deteriorates the quality of social life. In addition, similar signals are noticed with different physiological changes and symptoms. Thus, the subjective analysis makes the measurement onus and prone to error, and it hinders many quantitative information like disease state, duration etc.

Quantitative methods can overcome the shortcomings of qualitative methods. Thus, there is a dire need of a systematic approach to create a trade-off between reliability and consistency which makes multi-view information fusion crucial. It is worth noting that in supervised learning, the learning patterns are specified as members of a pre-defined set. Nevertheless, in practice, it is difficult to accurately label the signals that correctly represent the class of disease. Besides, the nature of changes and exact data distribution belonging to various disease groups, are still unknown. Nature of signals also depends on experimental setting, age of subject and subgroup of disease. In such cases, feature fusion based multi-view learning would be more relevant and promises to provide a more robust solution for classification tasks.

Real-world inference systems rely on the evidence of multiple signals corresponding to the same phenomena [1], e.g., audio-visual and EMG analysis. Each signal poses a different degree of uncertainty and does not have the same level of confidence, reliability and information quality. Therefore, use of large data becomes essential for deriving efficient inference system for complete knowledge of underlying phenomena [2,3]. Major challenges lie in effective management of large data by making an appropriate sense. Therefore, a proactive data-driven strategy that can provide a key probe that outweighs the challenges, is needed. It could enhance the uniqueness, interpretability, and robustness of models.

This thesis investigates the challenges of building feature fusion based learning system for biomedical signal analysis. It aims to utilize large-volume signal information in analyzing biomedical signals. The contribution of thesis includes design, implementation, and evaluation of a multi-view feature fusion based learning system. The rest of this chapter is organized as follows. We first address the motivation and challenges that promote to carry out this work in Section 1.2. Then, we review the state-of-the-art technique, information fusion, information fusion in machine, multi-view learning models and quantitative EMG support systems in Section 1.3-1.3.5. Furthermore, the challenges inherent to various state-of-the-art approaches as well as high-level look at proposed design and contributions, are addressed. The chapter concludes with thesis outlines in Section 1.6.

1.2 Motivation and Challenges

A rising number of data capturing devices and studies tremendously grow data traffic in the medical profession. Such data traffic is obvious and clinically important in addition to many engineering domains. Thereby, an efficient use of information gathered from large-volume data to develop support system and to expedite large-scale diagnosis research, is an indispensable provision. Promising potential of large volume data analysis as seen in [2–4] has motivated us to carry out medical diagnosis researches. Motivations of large-scale data analysis are listed as follows:

i). Real-world measurements often involve a large number of diverse sensor signals. Each sensor signal does not carry same level of information and also contains a different degree of uncertainty. Therefore, use of large data to extract suitable features for support systems to be applied healthcare, is essential [2]. Such strategy of covering wide varieties of data or signals has the capability to provide finer details of undertaking phenomena.

ii). For signal-based inference models, challenges lie on the learning framework since the performance highly depends on appropriate utilization of available signal information rather than the choice of classifier. Failure occurs when a learning strategy is unable to cope the possible information. Such failures negatively impact the society-support-systems, medical research, multimedia, large-scale diagnosis. Furthermore, they cause delays, cost overruns, failures on deployment, and even abandonment of projects. Therefore, most inference models focus on learning framework so as enhance the reliability as well as interpretability of model.

iii). Sort of information and large dimensionality are major concerns in high dimensional learning. Therefore, integration of data or information using feature fusion technique becomes essential. It intelligently integrates the multiple modality information to provide the decision maker with the best possible information on the likelihood. Multi-view feature fusion, a fusion-based learning system that aims to be used by a wide range of medical applications, has been designed that explicitly incorporates these challenges. Feature fusion based model simplifies the model interpretation and enhances the robustness of the algorithm for real-time medical applications. Thus, it can curtails the bridge between medical and engineering profession.

The fundamental lesson learned from early work, practical necessity and documented projects is that it is persuasive to maintain the consistency, reliability, and performance of learning strategy for practical values. A proactive approach to such management system has the benefit of anticipating probable performance problems. The goal is to develop a learning algorithm that meets overall objectives to have optimal performances. Suitable and reliable algorithm integration in healtcare devices could corroborate the users for early diagnosis and routine check-up without intervention of a physician. Thus, development of an efficient algorithm becomes part and parcel of engineered furnishings.

A number of challenges are still exist in many diagnosis methods. Over the last decade, a large number of methods have been reported in the bio-medical field. The primary objective of most of the learning techniques is to achieve high performance. However, many research quires such as reason of high performance, selection of particular features searched and assumptions etc., have not been properly explored. However, a few methods meet the necessity and hence, are practically implementable. The major challenges are listed as follows:

i). One of major challenge is involvement of constraints or assumptions which make them unsuitable for practical implementation. It further create the bridge between medical and engineering profession. Also, some assumptions are unrealistic in nature.

ii). In developing support algorithm, it is fundamental to understand whether performance requirements are fulfilled or not. Unfulfillment of requirements leads to absurd consequences. The early development phases heavily affect the quality of final decision and wrong steps at early phases trivialize the system or imply an expensive rework. Therefore, challenge is to find an appropriate learning strategy that can provide true interaction between model and input spaces of available and minimize the feature variances in learning phase.

iii). The utility and reliability of portable healthcare devices that rely on efficient program embedded promote to employ an advanced signal processing and diagnostic algorithm for automatic finding the problems and detecting alarming health trends to the users so as to inform medical professionals for further supports. Therefore, intelligent machine learning systems have high demand so as to make reliable Health Monitoring Devices (HMD). It enables recordings, customizations and hence provides unique information to identify alarming trends in the health status of the users. With such efficient algorithms, users can assess the health state by fusing information using intelligent inferences and learning. So, building of such an efficient learning is one of the major challenges of undertaken project.

The systematic research over large-scale populations is the key direction towards acquiring many vital consequences which tell about the models - their behaviors and modifications needed to adopt a practical design. Thus, it is reasonable to accept the above-mentioned consequences. Evidences derived based on large-scale analysis are acceptable only when the model validation and interpretation in the learning phase with already known database have a worthwhile effect. In case of failure, either the influential parameters or the overall system needs to be updated so as to get feasible outcomes to make the model substantial in the large-scale study. It is, therefore, important to understand how effective the learning strategy and variation of results on best available evidence are. Best practice for deriving potential techniques systematically is to utilize reliable information via an appropriate formulation. Subsequent validations of models require careful phased approach with a key focus on ingredients involving uncertainties in the scheme so as to get an exploratory and decisive estimation. Besides, it should be dispersed as widely and persuasively as possible with further research to assist and monitor the process of implementation. Support system having these abilities will be of great values in large-scale studies and diagnosis research. Such key learning approaches have the potential benefit over subjective error-prone analysis. Addressing such key issues and taking them as challenges, this study has been undertaken to explore in details.

1.3 State-of-the-art techniques

This section outlines a brief overview of information fusion (IF) and then presents the state-of-the-art approaches dealing with IF in computer vision and machine learning, and quantitative EMG analysis. Following the approaches used in quantitative methods and multi-views learning for which IF would be useful, we introduce the basic system design of fusion-based learning algorithm that helps the readers to follow the design proposed in subsequent chapters. In particular, we identified four principal categories of state-of-art-techniques:

i). Information fusion (IF): It utilizes strategies to cope multiple views or multiple modalities information to enhance system performance. The term multiple modalities indicate the sources that provide same or different type of signals for a given phenomena.

ii). IF in machine learning applications.

iii). Multi-view learning models that explore the problems dealing with multiple information embedding and iv). Quantitative EMG analysis.

1.3.1 Information fusion (IF)

Information fusion indicates the concatenation or integration of multiple input features, referred to as multi-view (MV) obtained same or different modalities data [5]. Albeit different views reflect their typical behavior, learning from all views helps in characterizing the object efficiently. Objective of IF is to cope inherent information of various views by using optimization technique. For uni-modal inference system, such as EMG analysis, MVs are generated from multiple sets of signals of a particular process or disorder in order to cover more diverse information. Noises often contaminate modality data. IF techniques help removing unwanted components from feature data sets before extracting usable features. Thus, it reduces the feature dimensionality without loosing the generalizaation ability of feature space. As a result, it has garnered attention in machine learning applications. It is divided into three categories as follow:

i). Data or feature level fusion. Feature level fusion is believed to be more effective for learning models that highly depends on modality data [1,6].

ii). Model fusion. It is required while modalities convey different characteristics information.

iii). Decision level fusion. For multimodal analysis, e.g., audio-visual fusion [7], where modalities carry complementary information, decision level fusion at final stage combines each modality output to make an appropriate sense.

1.3.1.1 Feature level fusion

A learning system to be useful for larger applications must be able to present an easy-to-use interface and shoulder a large fraction of the overall performance management burden (e.g., large scale integration, low variance, balancing, etc.). Best practice is the profound understanding and utilization of resources information in an efficient manner. Integration of multiple features in the initial framework creates high-order features that enhances the computational complexity. So, precise portrait of information (e.g., wavelet, statistical features etc.,) in low dimensional space which comprehensively depicts the underlying information, is essential. Classical data integration methods directly integrate data result in high order feature space. In contrast to that, feature level fusion enables interaction of multi-view input features, and extracts usable features for learning tasks [8]. Many applications include multimodal fusion, audiovisual fusion [7], image fusion [9], multisensor data fusion [10] and structured data fusion [11]. Learning model with well-defined coordinate system has sound feasibility and this arises due to its improved confidence, robustness and reliability, and high discrimination power [12].

1.3.1.2 Model fusion

Model-based fusion, attributed to a global model to achieve the final decision, combines multiple model's features. Many real-world inference systems require information from multiple modalities sources for proper interpretation. For example, brain imaging analysis requires multiple modalities-functional magnetic resonance imaging (fMRI), electroencephalogram (EEG) and structural magnetic resonance imaging (sMRI) [13]. Combined models are artificial intelligent techniques. Each sub-model processes individual modality information using feature fusion technique, which is then merged into global unit (i.e., high-level fusion), for an appropriate senses.

1.3.1.3 Decision fusion

High-level fusion, i.e., decision-level fusion is the fusion of outcomes from various processing stages. It takes decisions or symbolic representations as an input of various sub-models and combines them to obtain a more confident global decision. For example, Bayesian approach for binary event detection proposed by Krishnamachari in [14] that detects and corrects measurement faults. Objective is to take advantage of the redundancy of a set of independent classifiers to achieve higher robustness by combining their results [15]. Thus, overall system performances rely on the resources utilized by the sub-models. Feature fusion framework involve in each sub-module to make full utilization of individual modality information. This way, it spans the information space to improve the performance management burden of models.

1.3.2 Information fusion in machine learning

1.3.2.1 Feature projection techniques

The goal of feature projection is the transformation of feature space to a well-defined coordinate system. From the projected space, the low order latent features that comprehensively represents the underlying phenomena are estimated using a given optimization criteria. Feature projections such as principal components analysis (PCA) [16, 18, 19] well approximates the distribution of original space without considering the class information. PCA provides good feature clustering assuming variance in data [20]. However, large input space alters the feature clustering to a complex pattern which will degrade the performance of the PCA. It is worth mentioning that issue encountered in PCA can be efficiently handled by linear discriminant analysis (LDA) [21, 22] incorporating class-structure information. In other words, it finds the optimum decision surface by enhancing the separation margin among multi-group features exploiting the class information which makes it more popular in learning applications. Furthermore, some generalized versions-uncorrelated LDA [23], non-linear LDA (NLDA), non-linear unsupervised self-organizing feature map (SOFM) [16] are also popular in literature. Uncorrelated LDA (ULDA) incorporates the optimization criterion like LDA and generalized singular value decomposition, which mitigates the issues related to singularity [24]. In PCA [16], the transformation coordinates of projected features are uncorrelated and the maximum variance of the original feature space is preserved by a few coordinates. First, $\Sigma \in \mathbb{R}^{n\times n}$ is formulated from the feature space and then, following the steps, a $W \in \mathbb{R}^{n\times k}$ is formed, where columns consist of k eigenvectors corresponding to k largest eigenvalues from the covariance matrix. According to $y = W^T P$, original feature space P is, then, transformed to k-subspace. NLDA [25] involves a nonlinear transformation of a multilayer perceptron and supervised learning. SOFM involves synaptic weight vectors adjustment based on their similarity to the input pattern without class information, for which it sometimes transforms the input patterns for different classes into the same cloud, causing a low recognition rate.

1.3.3 Application of information fusion

1.3.3.1 Canonical Correlation analysis (CCA)

Canonical correlation analysis (CCA) was first developed by Hotelling in 1936. Since then, it is widely used in many engineering applications. It provides a unified framework that transforms the input features to a well-defined coordinate system and fuses the obtained low order features using parallel and serial fusion technique to derive single form of discriminant vector for pattern classification [6]. It has high proficiency in extracting unique feature representation depending on the degree of proximity between two feature sets extracted from the same or different objects [26]. It searches two set of vectors, one for each input variable, such that they are maximally correlated in mapping space. Low order features inherently carry unique information to characterize the object [27] and it has been successfully applied for handwritten character [9, 28, 29], face image [5, 30–33] recognition tasks, EMG signal analysis, filter etc. Besides, other applications of CCA and its extensions are also of note.

Application of CCA and its extensions are briefly summarized herein. In [5], a CCA-based feature extraction and feature fusion scheme was adopted for arabic numerals, yale face and handwritten recognition. In this scheme, taking two groups of feature vectors with the same pattern and measuring the correlation function between them, highly correlated features were extracted to represent discriminant patterns for learning. Shen *et al.* [28] introduced an orthogonal multi-CCA (MCCA) based on fractional-order dimension reduction technique to avoid small size problem in MCCA. The authors conducted test with *UCI** multiple feature data set and CENPARMI handwritten Arabic

*C. Blake, C.J. Merz, UCI Repository of Machine Learning Databases [http:// www.ics.uci.edu/ mlearn/MLRepository.html]. Department of Information and Computer Science, University numerals database to validate the model. In [30], a cascaded generalized CCA was introduced for face recognition using information about images at the various level of resolutions to enhance the information content of the features. To fulfill the objective, high resolution Gabor-feature hallucination of face images were estimated directly by using local linear regression. It was extended to include low resolution features too. Extracted features were fused for recognition and reported promising results. Sun *et al.* [32] proposed a generalized canonical projective vector intending to enhance the performance. It utilized the class information of samples so as to obtain more discriminant features. The validated result using Concordia University CENPARMI handwritten arabian numeral database indicate efficacy of the model.

A multi-set integrated CCA framework was introduced in [29]. It establishes a discriminant correlation function of multi-group variables based on generalized correlation coefficient to comprehensively depict the integrated correlation among multiple features. The extracted multiset integrated features were fused through two given feature fusion schemes for subsequent estimation of discriminant vectors for handwritten pattern recognition. Peng *et al.* [31] employed a local discrimination-CCA which combines the local properties and discrimination between different classes in addition to the inherent inclusion of correlated features among the class variables. In addition to that, a kernel version of local discrimination-CCA was introduced to cope with nonlinear problems. In [28], a two-view semi-supervised CCA model was adopted in order to incorporate a new sparse representation-based label propagation algorithm to infer label information for unlabeled data. A dictionary was constructed for all labeled samples and using sparse representation technique, reconstruction coefficients of unlabeled samples were estimated. Using labeled sample information, label information of unlabeled samples were estimated, followed by label matrices of all samples and probabilistic within-class scatter matrices in each view. This method aimed to maximize the correlations between samples of the same class from cross-views while minimizing within-class variations in the compact space of each view simultaneously. The method was further extended to analyze multiple (more than two) views for classification task. Allan [34] employed CCA and MCCA for data fusion, multi-source and multi-temporal exploratory image data. Many other applications of CCA and its variants include facial expression recognition [35], position estimation of robots [36], parameter estimation of posture [37], data regression analysis [38], image texture analysis [39], image retrieval [40], content-based text mining [41] and asymptotic convergence of the functions [42].

In addition, CCA and MCCA are widely used in medical as well as signal processing domains. For instance, in [13] the authors proposed CCA and MCCA-based data fusion schemes to study the brain imaging modalities, fMRI and sMRI and inter-subject covariance of fMRI, sMRI and electroencephalography (EEG). The illustrated results indicate the suitability of data fusion scheme for multi-modalities applications. Sargin *et al.* [7] explored the effectiveness of early fusion technique and introduced CCA-based

of California, Irvine, CA, vol. 55, 1998 multimodal fusion strategy for open-set speaker identification using a combination of early and late integration of speech and lip texture features. The developed methodology derived a high precision synchronization of the speech and lip features. From the thorough analysis and results, the authors concluded that this scheme can be applied to many real-world fusion of any pair of modalities, which can be modeled as a synchronization of correlated and uncorrelated components. In [43], a CCA-based frequency recognition scheme was addressed. Method was then applied to analyze frequency components of steady-state-evoked potential in EEG. It employed CCA optimization to extract dominated frequency features of evoked potential for brain computer interface (BCI) based detection and recognition scheme. The method has high recognition ability as compared to the traditional power spectral density analysis (PSDA) based recognition method. Usually, traditional filters fail in extracting the noise or uninterested components, in case when they fall within the region of interest signal components. More specifically, it fails in case of noise-frequency overlapping with that of signal frequency. In this context, above method is suitable for analysis and outweigh the challenges. De-Clercq *et al.* [44] developed a CCA-blind source separation technique to separate muscle artifacts from the EEG signal and reported promising results. In [27], CCA was employed for multi-user myoelectric interfaces.

Development of learning systems relies on the choice of input patterns, feature projection, dimension reduction and subsequent evaluation of generalized discriminant representation. Choice of patterns and their thorough combinations are primarily important so as to get all complementary information and the characteristic nature of inputs. The generalized criteria enable the visualization of high-dimensional pattern vectors in low dimension and allow to evaluate relevant features for selection of classifier with the best performance. Furthermore, it greatly improves computational efficiency and makes real-time implementation possible. Major advantage of CCA based learning is that it transforms the original feature space to well-defined coordinate system and measures the mutual information between same modality views. As a consequence, it provides a better feature selection approach, better insights and finer details of the problems.

1.3.4 Application of multi-view learning model

Recently, there has been growing interest in multi-view learning (MVL) for various applications including computer vision and pattern recognition applications [9, 45–48]. As mentioned earlier, most of the real world bio-medical inference system relies on multiple observations or multi-views captured during the experiment. This is due to the typical nature of nonstationary signal. In multi-view learning models, the same object is represented by different views or data with same or different statistical nature. Therefore, an effective use of multi-view feature reduces the variance i.e., error in the feature registration and improves model learning ability, performance and reliable decision outputs.

Notable efforts in the recent past include multiview data learning [49], classification [50], feature selection [51], and clustering [47, 52]. Nonetheless, progress over dimension reduction strategy in multi-view learning is not much noticeable, whereas it has already made successes in multimedia applications [53], e.g., image retrieval and video annotation. Xia *et al.* [45] proposed multi-view network that encodes different information in different ways to have meaningful physical embedding for image retrieval, video annotation, and document clustering. Zhou *et al.* [47] developed multi-view spectral clustering via generalizing the normalized cut from a single view to multiple views and then, multiview spectral clustering was developed for web classification. Liu *et al.* [48] adopted an inherent structure based multi-view model for classification of Alzheimer's disease and its prodromal stage (i.e., mild cognitive impairment or MCI). In this approach, multiple feature representations were generated for subjects using multiple signals and then subjects within a specific class into several sub-classes (i.e., clusters) in each view space were clustered. Multi-task features were extracted encoding sub-classes with unique codes by considering both their original class information and their own distribution information. The model was trained with ensemble-support vector machine (SVM) for classification and reported promising results. Inspiring from the success and multifarious advantages, we propose a framework to explore the effectiveness in EMG signal analysis.

1.3.5 Quantitative EMG support methods

1.3.5.1 EMG signal

EMG signal is the electrical manifestation of the neuromuscular activation of skeleton muscle [54]. It carries significant information for diagnosis of neuromuscular disorders which can affect motor units (MUs) [23, 55, 56]. Fig. 1-1 shows typical one dimensional EMG signal pattern and motor unit action potential (MUAP). Due to disorders, significant changes are observed in the morphology and physiology of MUs and MUAP carrying the signature of disorders [57]. The physiological changes of muscles due to disorders, e.g., ALS and myopathy can be seen from Fig. 1-2. Therefore, quantitative assessment employing EMG signals or MUAPs is essential for early diagnosis of diseases.

1.3.5.2 Neuromuscular disorder

Neuromuscular disorder affect an overwhelming number of wold's population and deteriorates the quality of life. It refer to disorders that affect any part of the nerve and muscle. They vary according to the characteristics such as pattern of inheritance, origin of the genetic mutation,incidence, symptoms, age of onset, rate of progression, and prognosis. Table 1.1 outlines the summary of various neuromuscular disorders reported in literature. The abbreviations used herein are congenital myasthenic syndrome (CMS), neu-

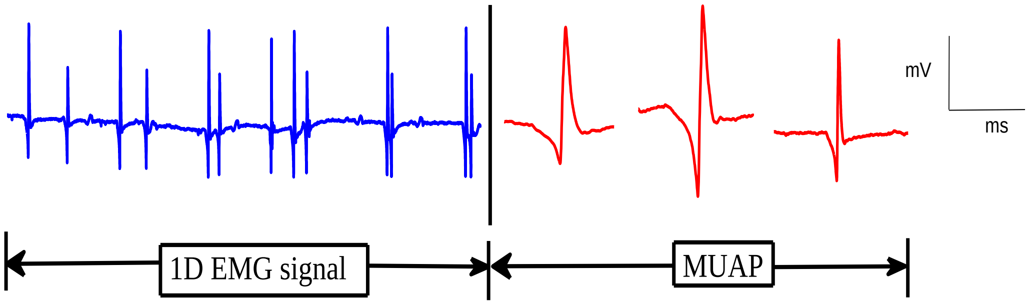

Fig. 1-1: One dimensional EMG signal and its corresponding dominated MUAPs. Note that EMG or MUAP amplitude fall in the range of 0-300 mV.

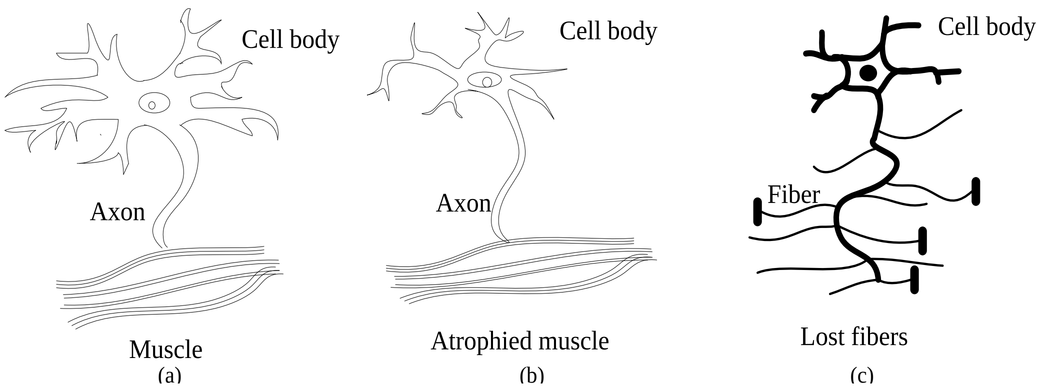

Fig. 1-2: Physiological changes associated with disorders. (a) normal (b) ALS and (c) myopathy subject. Loss of muscle mass and fibers are observed due to ALS and myopathy [55,56].

romuscular junction (NJ), peripheral neuropathy (PN), GuillainBarr syndrome (GBS), peripheral nervous system (PNS), muscular dystrophy (MD), spinal muscular atrophy (SMA), myasthenia gravis (MG), mitochondrial myopathies (MM), cramp fasciculation syndrome (CFS), duchenne muscular dystrophy (DMD), inclusion body myositis (IBM), neuromyotonia (NMT), LambertEaton myasthenic syndrome (LEMS), congenital myopathy (CM) etc.

1.3.5.3 EMG Features

Typically neurophysiologist assesses a large number of signals or MUAPs (at least 20) and estimates a set parameters (duration, amplitude, area, number of phases and number of turns) to compare with age-matched normal references derived from the available database. This assessment is subjective, error prone, time consuming and requires experienced physician. It fails to provide many important insights and quick diagnosis in urgent clinical cases [58,59]. Furthermore, it precludes the large-scale population studies in diagnosis research. Development of EMG precision decomposition (EMGPAD) that incorporates automatic decomposition of MUAPs from EMG, and EMGLAB (Fig. 1-3) have made significant improvement in this era [54]. Challenges are still there and many

Table 1.1: Summary of neuromuscular disorders, causes and effects [17, 56].

Disorder	Cause	Effect
ALS	Death of neurons	Stiff muscles, twitching Lost of muscle size
CMS	Defects at NJ	Loss of mobility respiratory problems
CM	Skeletal muscle fibres	Muscular weakness, hypotonia
PN	Systemic diseases, vitamin deficiency medication, traumatic injury excessive alcohol consumption	sensation, movement gland
GBS	Damage of PNS	Rapid-onset muscle weakness
MD	Breakdown of skeletal muscles	Muscle weakness
Myotonic dystrophy	Lost muscle functions	Worsening muscle loss weakness
SMA	Loss of lower motor neurons	Progressive muscle wasting
MG	Skeletal muscle weakness	Eyes, face, swallowing
Myopathy	Loss of fibers	Inability to move, walks
CFS	Damage of PNS	Fasciculations, cramps, pain, fatigue, muscle stiffness
DMD	High muscle mass	Muscle weakness
HSP	Gait disorder	Progressive stiffness contraction in the lower limbs
IBM	Lost of muscles	Progressive weakness, wasting
NMT	Peripheral nerve hyperexcitability	Spontaneous muscular activity
LEMS	Autoimmune reaction	Muscle weakness of the limbs
Neurogenic	Damage of PNS	Stroke, dementia, Parkinsons disease, tumor etc.

research queries remain unanswered. As a result, extensive research efforts have been given to provide more sophisticated diagnosis probe that helps depth understanding of underlying process and subsequent diagnosis.

Many support system employ data-driven models to improve the diagnosis process. Various methods employ direct features of EMG signals or its extracted MUAP features to diagnose neuromuscular disorders like *amyotrophic lateral sclerosis* (ALS) [55], myopathy [57] etc. A comprehensive review [60] outlines the methodologies and their features in steady improvement of classification performance. Fig. 1-4 shows the tree diagram of features. Time domain or MUAP morphological features as shown in Fig. 1-5 employed in most of earlier methods are listed below [57]. In the context of achieving promising performance, various methods employed direct EMG features (e.g., zero crossing) [61, 62], MUAP features (i.e., duration, amplitude etc.) [63–65], frequency features [66], wavelet-feature [67–69], autoregressive coefficient (AR) [70–72], time-domain (TD) [73], AR+RMS [74, 75], multiscale PCA features [76], empirical mode decomposition (EMD) and intrinsic mode functions [23] extracting from specific diagnosed signals with special emphasis on classification models. Some approaches utilize statistical mea-

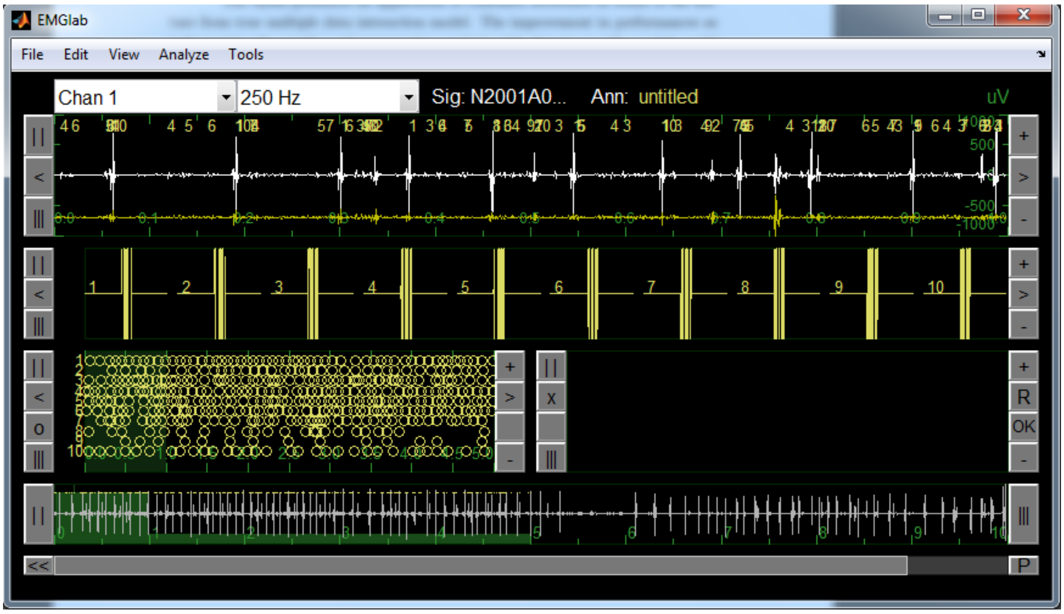

Fig. 1-3: EMGLAB with single channel EMG pattern filtered at 250 Hz.

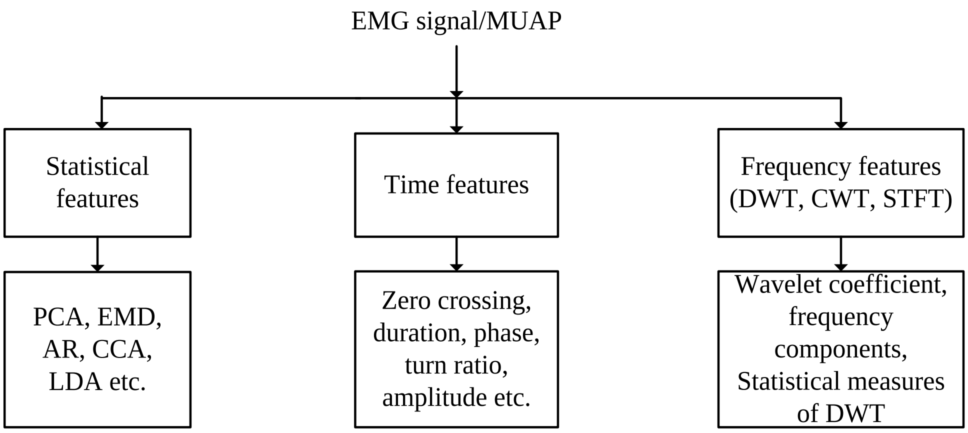

Fig. 1-4: Tree diagram of feature used in quantitative EMG support methods.

sures of feature vectors for classification [77]. Various models include adaptive neuro-fuzzy inference system (ANFIS), SVM, particle swarm optimization (PSO)-SVM [78], ensemble-SVM [79] and radial bias function (RF) [80].

i). Rise Time: Time between the initial positive to the next negative peak within the main spike.

ii). The time between start and end point of a MUAP

iii). Spike Duration: The time between the first to the last positive peak.

iv). Peak-to-Peak Amplitude: Amplitude difference between the minimum positive and the maximum negative peak.

v). Area: Rectified MUAPs integrated over the calculated duration.

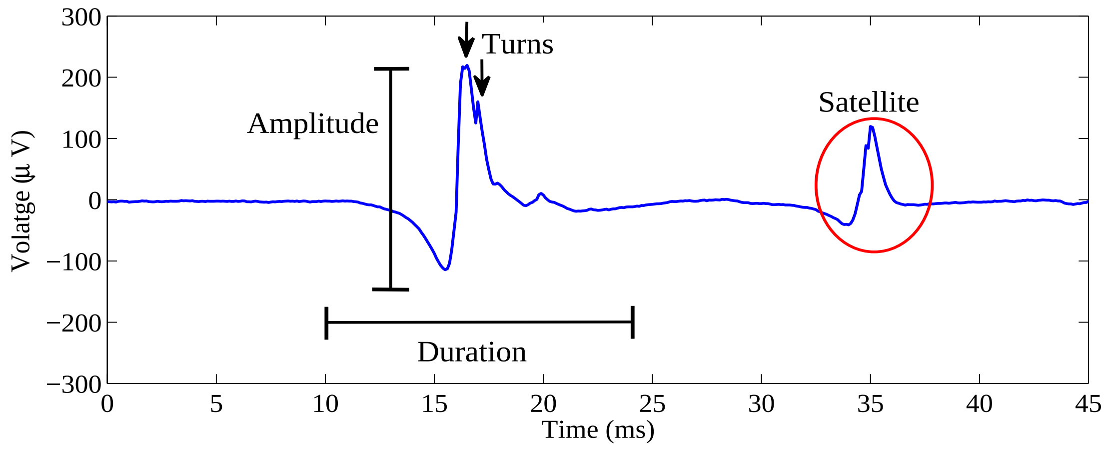

Fig. 1-5: MUAP morphology and features such as duration (ms), amplitude (μV), turns etc.

vi). Phases: The number of baseline crossings where amplitude exceeds $\pm$ 25 μ V, plus one.

vii). Turns: The number of positive and negative peaks where the difference from the preceding and following turn exceeds 25 μ V.

viii). Thickness: The ratio of the area to the peak-to-peak amplitude.

ix). Peak-to-Peak Samples Number: Total number of samples between the minimum positive and the maximum negative peak.

1.3.6 Classification paradigm

Quantitative methods that employ various features as outlined in the section 1.3.5.3 are briefly summarized herein. For instance, Subasi *et al.* proposed a discrete wavelet transform (DWT) based neuro-fuzzy inference system to classify neurogenic and myopathic disorders. This method employed the dataset of 27 subjects with multiple signals over each subject collected in Neurology Department of University of Gaziantep. Doulah *et al.* [58] introduced the method based on DWT and statistical features elicited from dominated MUAP to classify normal, myopathy and ALS using k-nearest neighbors (k-nn). The performance was investigated over online database [54] that includes 8 ALS, 7 myopathy and 10 normal subjects. Despite significant performance, this method requires a priori of morphological patterns to evaluate dominated MUAPs for feature extraction. Furthermore, DWT based method often requires human intervention and the task of matching wavelet coefficient from each level with the specific MUAP is difficult. Also, use of specific wavelet coefficients from each level may not be feasible for all MUAPs [23]. In addition to that, various models have pros and cons [23].

The author in [63] reported a parameter-based algorithm that facilitates the combination of automatic MUAP feature extraction and artificial neural network (ANN) for diagnosis of neuromuscular disorders. It employed MUAP features and reported an accuracy of 80% over dataset of 14 normals, 16 motor neuron disease and 14 myopathy subjects. In [64], the authors adopted a two-stage approach based on EMG decomposition and clustering techniques. It first decomposes signals and evaluates class specific template MUAPs using an averaging method. In classification stage, it used radial basis neural network to filter out normal MUAPs and then, pathological templates were further classified using decision tree to into neuropathic or myopathic. It reported an accuracy of 94.17% over the database of 62 subjects (20 normal, 20 myopathy and 22 motor neuron disease subjects). In [65], another MUAP-based approach using ANN to an early and accurate diagnosis of diseases was proposed. The model validation was based on 800 MUAPs from 12 normal subjects, 15 subjects suffering from motor neuron disease and 13 subjects of myopathy.

Frequency-domain features are promising in carrying inherent characteristics of signals of the process of interest, as indicated by the application's success. As mentioned in earlier discussion, independent time or frequency, time-frequency are widely used and the studies argued the potential benefit in applying them in real-time problems. Many reported methods employed combination of time and frequency features in order to improve the generalization ability of the feature space. These features are statistically independent and complementary in nature. So, combined feature space adds intrinsic information about the undertaken process. Transformation used in various methods such as DWT reveals the occurrence of the particular frequency band in a specific time interval, which is one of the major concerns for real-time signal analysis. Wavelet transformation provides good resolutions in time-frequency for high and low-frequency components. It basically decomposes the signal into high and low-frequency components and thus, it acts like a filter [81]. One can easily choose appropriate filtered components depending on one's requirements using domain-specific or ground subjective knowledge for analysis. The use of wavelet function requires a trade-off in fixed frequency scale and sampling frequency. Therefore, important task in such methods is to find out proper wavelet function that accurately fits the user signal and which determines the high and lowpass filter coefficients. However, with proper procedure and repeated investigations, users can avoid the shortcomings encountered in such methods. Maitrot *et al.* [82] developed an approach for such selection based on feature space and a distance measure. Many typical examples which achieved the success include [69, 83–85].

WT-IP EMG [83], proposed by Abel *et al.*, is a singularity characteristics measures for IP EMG which can divulge more diagnostic information than their constituents MUAPs. The method evaluated singularity and irregular structure through WT technique to depict the information contents in the signals and to measure singularity characteristics of EMG IP using Lipschitz exponent. In testing phase, the method was applied over database of 11 normal, 26 myopathic and 20 neuropathic subjects, and found that WT-IP EMG is superior to the traditional "turns analysis methods". In [69], a scalograms-based method using Symlet4 wavelet for disease diagnosis. It involved the isolation of single MUAP and computation of scalograms for each subject. A single parameter based decision strategy (i.e., wavelet index) was used to identify the myogenic, neurogenic or normal subjects. Other such schemes are also seen in [68, 85].

Subasi [70] introduced a medical decision support system by combining DWT and fuzzy-SVMs. Estimating the time-frequency representations of various classes, obtained low order statistical features were embedded to the models. In [67], the usefulness of various feature extraction techniques for describing MUAP morphology were discussed and a soft computing approach was developed for signal classification, followed by comparison analysis with multilayer perceptron neural networks, dynamic fuzzy neural network and ANFIS. In addition to that, similar methods reported in literature include DWT+ANN [86], DWT+SVM [87], DWT+Decision [80], multi-scale-PCA [76], DWT+ESVM [79], which employed different feature extraction and concatenation techniques for discriminate feature representations. Besides, methods SOFM+LVQ [88] and probabilistic neural network [89], PCA and PNN [90], Fuzzy logic [91] are also noteworthy in this study. For instance, in [57], Kamali *et al.* estimated various feature representations from MUAPs based on prior knowledge, which were further truncated using statistical measures. Despite simplicity due to low order feature, this method may not be feasible for all applications with wide varieties of signals since it is seen that in many cases statistical features of specific group overlap with others.

Apart from inherent limitations of models, challenges lie on 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]. Failure occurs when a learning strategy is unable to cope the possible information within the framework. Although most of the aforementioned methods reported positive results, feature extraction stages which involve signal selection mechanism, reliable feature extraction strategies, were less exposed. Thus, there may be the possibility of feature biasing which will result in loss of reliability of the algorithms for clinical practice and health alarming systems. Generally existing methods usually focus on utilizing the different combination of feature vectors extracted from specific diagnosed templates associated with studied groups. Subsequently, such approaches simplify the problems and learning models. However, it degrades the generalization ability of models and reliability despite their promising results over limited database. In supervised learning models, it is difficult to accurately label the biomedical signal which correctly represents the disease class [48]. The nature of signals depends on disease profiles, nature of force recruit to muscle and measurement setup. As such, features elicited from specific predefined signal might fail in providing the exact class information to which it belongs. As a consequence, multi-view probing technique where specific learning group features are evaluated from multiple input features associated with same study pattern, is considered as a remedial measure. It could provide a more robust solution for disease classification and other applications [93–97].

1.4 Human computer interaction and outlook of feature fusion

Human-machine-interaction (HMI) is heavily inspired by natural phenomena. In HMI, design modalities make the analysis natural, efficient and as intuitive as possible [98]. Basic idea of designs is to integrate multiple interactions and interpretations. Therefore, it aims to make strategy so as to actively shared modality profiles or multiple information sources for decision module rather than focusing one of them as active component. In doing so, human users play active role in choosing out appropriate model or design for true interaction or fusion among modalities. However, practical reliability of fusion procedure is relies on the the ability to adapt with technological combination of data streams and predict the outputs. This study therefore, has provided an efficient HMI design for possible implementations for alarming health trends or classification disease offline or online and to promote large-scale diagnosis research.

Decision models predict unknown class patterns relying on the information preserved during training phase. The complexity and largely unknown behavior of the underlying phenomena, including various perplexes (see, section 4.2), specific research queries turn out the high level data fusion more useful in real scenario. Introducing relevant issues, the feature fusion-based data-driven models are devised and examined in both chapter 3 and 4. The hypothesis of our design is to focus address and depict the integrity of parallel information processing pipelines i.e., combining statistics from two independent domains, followed by decision-making steps. It allows the features to fully interact and inform each other for making up the global fusion with prior assignment of symmetrical role to all features. Further, it restricts the degrees of freedom and constraints. The design adopted in this study involves two-stages. In the first stage, feature vectors are computed using a certain criteria function and In the second stage, feature vectors are fused using different criteria. A strategy that merges the two, and thus, this *data compression* is expected to better exploit the whole data. It has been shown to be successful in previous as well as in this analysis.

The modular outlook on feature information fusion offers beneficial provisions. A major challenge in feature fusion is the complexity due to high degrees of freedom. The modular technique responses this challenge by reformulating the problems in a well-defined way in terms of small number of components and constraints that can be separately optimized, analyzed and coded. Secondly, it factorizes the problem into smaller stand-alone elements and then combines to give a new emergent look up that promotes deriving models, algorithm and device implementation [99,100]. Thirdly, computational challenges in large-scale data analysis are easily adapted. Fourthly, it removes the bottlenecks and provides straightforward exploratory approach to test and compare alternatives with minimal effort. Thus, it is easily accessible and allows easy diagnosis tool with wide setting adjustment facilities.

Appropriate and simplified models can minimize the bridge between-in and out professions. Such models are often preferred in most of scientific domains including medical profession. Before implementation in real-world applications, model validation is indispensable for practical values and must be answerable to important research queries such as the best achievable errors with minimum fluctuation (i.e, SD) over large-set repeated measurements for specific task, goal, and model. The reliability and practical usefulness of models that are characterized by physically, meaningful, unique and easily interpretable are also commensurable issues. In addition, some new queries -"success", "error", and "optimality" of model while combining heterogeneous modality signals with uncertainty are also important. Further, target and criteria function, evaluation of task and figure of merit that inform how to exploit the advantages of each modality without suffering from drawbacks w.r.t. the other modalities are to be explored. In real scenario, it is difficult to fulfill all the criteria of success, however, a close approximation with simplified approaches towards a plausible solution for the problem are acceptable. A few of them can be made to trade-off to add value to specific applications. Our design approach comprehensively explores all the issues either directly or indirectly from the model outcomes and also answerable to many of crucial queries.

1.5 Objectives

This study aims at establishing a learning system that would be simple, reliable and flexible. Furthermore, it could also corroborate the maintenance of consistency in performance that allows applications success. We explore the possibility of efficient learning mechanism to enhance performance and expedite the diagnosis research. Accordingly, we develop a data-driven network employing feature fusion with aid of statistical models. The objectives of this research work are, thus, summarized as follows:

i). Investigations of various state-of-art-feature fusion and EMG diagnosis methods that include various signal processing techniques, feature extraction, fusion, brief exploitation of inherent shortcomings and possible improvements of methods by employing advance signal processing approaches for real-time implementations

ii). Development of high dimensional feature formulation, mathematical frameworks for embedding high dimensional heterogeneous EMG data and CCA based feature fusion algorithm for diagnosis of neuromuscular disorders using single channel intramuscular needle EMG signals.

iii). Development of a feature level fusion based data-driven model generalizing CCA, namely, (mCCA) model for classification of EMG signals and improvement of classification performance.

iv). Development of a novel multi-domain multi-view fusion based data-driven model generalizing discriminant correlation analysis (DCA).

In presenting new learning architecture, we propose various feature formulation strategies to efficiently cover wide-scale information of available input space and we also investigate the impacts of various formulations over performance. The algorithm is validated with various sets of real-time signals. In each stage, the algorithm is examined separately and compared with existing literature. Following the results, each section describes under what circumstances the proposed model is useful and how this design overcomes the shortcomings of previously reported methods. Finally, we also explain the cause of possible variation of models outcomes and how it affects for real-time process analysis and management.

1.6 Thesis outlines

In order to make this thesis accessible for readers, the work presented is divided into a number of chapters. Each chapter outlines the contribution provided to model evaluation. This is as follows:

Chapter 2 outlines an overview of EMG signal, the physiological process associated with and also it describes signal recording, acquisition setup and signal analysis. At the end of the chapter, the data sets used in this study are analyzed and summarized. In particular, the chapter is categorized as: (i) EMG signal, (ii) generation and structure of EMG that help understanding the morphological pattern of MUAP associated with signals, (iii) EMG signal analysis that gives the indication of morphological changes in pattern during pathological conditions, and (iv) datasets description.

Chapter 3 presents the proposed feature fusion based learning models for efficient use of high dimensional information. The mathematical frameworks developed for embedding high dimensional heterogeneous EMG data are explained and examined with real-time data. Basically, this chapter outlines two proposed feature fusion based decision support schemes for EMG classification problem. At the beginning, the chapter focuses on two independent feature selection and fusion mechanisms, followed by classification and discussions. Finally, it addresses the integrity of each approach and compares with the state-of-the-art methods.

Chapter 4 presents the work which aims at establishing a novel multi-domain multi-view fusion based data-driven model generalizing discriminant correlation analysis (DCA), namely MVDCA. The model is more efficient and capable of incorporating the class association in the feature sets, unlike the method presented in Chapter 3. It has the ability to decorrelate feature vectors which are common to different classes within each feature sets. The chapter address many shortcomings in current approaches-class-structure or class association, complexity, performance, reliability, adaption to noise and signal processing method and discussed in the context of our approach to overcome these limitations. The model first attempts to incorporate the class structure into the feature fusion so as to get comprehensive representations of class information from multi-view vectors and then it tediously examines and explores all necessary issues step by step. The efficacy of this generalized cascade model is discussed in analyzing the results. From the consequences, it is believed that this scheme will also be suitable for other signal analysis-EEG, ECG, and multi-modality signal. This Chapter elaborates well-defined framework to gain benefit over wide-scale learning scenario. Extensive experimental investigations demonstrate the effectiveness of adopted method that outperforms many state-of-the-art methods. At the end, this chapter also explores the variations of results in terms of various performance markers under different feature formulation strategies with different classifiers. It gives brief overview of the performance enhancement and selection of suitable strategy for real-time applications.

Chapter 5 provides the conclusions and future direction of research. It summarizes the overall work presented in this thesis with a special emphasis on feature fusion strategy for uni-model biomedical signal analysis. It further focuses on the possible way for analyzing the multi-modal signal analysis. Finally, the chapter concludes with the future direction of research briefly.