Abstract

Growing interest in biomedical research to support the early diagnosis, demands to develop compact health monitoring devices. Utility and reliability of such devices rely on the embedded algorithm so that appliance can manage the jobs without the intervention of clinicians. Keeping view on the present health situation of world's populations, early and precise diagnosis is becoming an urgent as well as emergent provision. Critical engineered changes could bring the balance to the society and enable prevention of most of the real-life problems. Thus, development of efficient algorithm employing an advanced signal processing for automatic finding the problems, is essential. Therefore, intelligent machine learning systems have high demand so as to make reliable health care appliances. With such efficient algorithms, users can assess the health state by fusing physiological or statistical information using intelligent inferences and learning.

Feature fusion have been garnering interest in pattern recognition community due to its robustness in capturing multiple view information through a given strategy to well-facilitate the classification tasks. It is an important data processing step and its adaption to derive efficient learning is essential. In medical domain, availability of multiple recordings which is due to typical nature of non-stationary signals or large modality profiles is inherent. Therefore, requirement of collective interpretation for clinical findings makes the conventional practice onus and prone to errors. Albeit number of successful efforts in this direction, yet many challenges are still there, which are to be resolved. Most importantly, feature biasing and curse of dimensionality are two major issues in the decision modules. Despite promising performances of various quantitative methods with efficient signal processing strategies, they often fail to answer many research queries such cause of performance enhancement, reliability, diversity in results etc. Most of the existing models utilize features elicited from specific distribution of profiles (i.e., signals) for learning result in diversity of model performances, which is presumably due to feature variance and its dimensionality. Therefore, in recent machine learning community, feature fusion researches have garnered attention. It enables to interact various signals and inform each other so as to extract usable features which are more discriminant and reliable for diagnosis applications.

In this research, we develop a data-driven network employing feature fusion with the aid of statistical models. This dissertation makes important contributions to the development of machine learning techniques, diagnosis analysis, large-volume population study and portable homecare application. Contributions of this dissertation can be

categorized to address four issues.

In the first contribution, the thesis provides an overview of electromyogram (EMG) signal, associated physiological process, signal recording, acquisition setup and signal analysis. Furthermore, it explore the details of real-time data sets. Then, we provide the solution for reducing the dimensionality of input feature space based on signal variability and stability measurements. The study aims at incorporating large-volume physiological or statistical feature information, which are to be fused and embedded to the learning models in order to diagnose normal and pathological patterns. Accordingly, a set of multiple input features or multi-view features (MV) are formulated using mathematical model in both direct feature space and discrete wavelet domains. These MVs would have most physiologically relevant information and energy contents of underlying phenomena. Subsequently, a subspace learning technique generalizing Canonical Correlation Analysis (CCA) is developed and, variability and stability analysis have been demonstrated. Based on this analysis, low order feature sets are evaluated and concatenated them using feature fusion strategies for discriminant evaluations which are subjected for statistical validations and classification task. Afterwards, the model performances are compared with various recent state-of-art EMG diagnosis methods.

In the second contribution, it formulates more generalized MV features by employing multiple signals associated with a particular study group. Unlike random signal selection strategy, this formulation uses multiple signals based on age specific subgroup of a given study group. It basically aims to cover more phenomenological information into the input features so as to derive more discriminant information than previously stated formulation. Similar to aforesaid formulation, it also incorporates multi-domain MV features in modeling and evaluating discriminant features using feature fusion model. In this analysis, the most efficient feature fusion model obtained from previous analysis, is adopted for finding discriminant features. Features are subjected to statistical test-one way analysis of variance (ANOVA) in order to find the most dominant features. These features are embedded to various classification models and learning performances are investigated in the context of various medical diagnosis markers including accuracy (Ac), specificity (sp) and sensitivity (Sn).

In this case, results are further compared with various state-of-the-art methods in two different ways. In first case, the results obtained using proposed algorithm is compared with the reported methods which shows the superiority of our scheme. Since reported methods employed different classification models as well as databases, in the second case, we have implemented their feature extraction techniques and feature structures over our database. The proposed model shows better performances that indicates high discriminate ability of our proposed feature fusion scheme.

In the third contribution, a multi-view information embedding scheme is developed using discriminant correlation analysis (DCA). DCA fusion framework contains similar theoretical formulation with the CCA with exception of feature discriminant function. Unlike previous contributions, the objective herein is to develop high dimensional multi-view model, namely, multi-view multi-domain-DCA (MVDCA) incorporating structure

information into the feature space simultaneously decorrelating between group features. MVDCA finds the multiple domain canonical variates summarizing the correlation structures among the multiple variables by linear transformations. In addition, optimizes the criteria function between features in multiple domain independently to obtain maximum overall correlation and more meaningful structure for variables corresponding to same phenomena. Accordingly, a well-define mathematical framework of MVDCA is developed through which multi-domain MV features are embedded so as to get comprehensive representations of class information from multi-view vectors and then tediously examines and explores all necessary issues step by step. Successful implementation of the algorithm to diagnose neuromuscular disorders and comparison with the state-of-the-art EMG and nonlinear industrial process diagnosis methods reveal the effectiveness of the MVDCA. Furthermore, in the context of performance evaluation, various shortcomings in traditional methods, as well as other important issues like dimensionality, reliability, easy accessibility and robustness for viable implementation are discussed thoroughly.

In the fourth contribution, we have analyzed the variations in results obtained from various data-driven approaches in multiple experiments to properly choose the models for suitable implication. It explores the ability of models' interactions with various natural phenomena evolving phenomenological characteristics. This section further addresses four key issues; 1) performance diversity and necessity, 2) the nature of inputs that to be embedded into the models, 3) impact of feature level fusion on the diversity in performances and 4) diversity in the choice of appropriate model.

Finally, the thesis concludes with description of methodologies, significant achievements, limitations and future direction of research. It basically analyzed the gain benefits of various aspects: multi-view feature creation, fusion strategy, theory, algorithm, real-time implementation comparison analysis.

Keywords: Information fusion. Feature fusion, canonical correlation analysis (CCA), multi-view multi-domain discriminant correlation analysis (MVDCA), Electromyography (EMG) and discreet wavelet