French Polytech network form for PhD Research Grants from the China Scholarship Council

This document describes one of the PhD subjects proposed by the French Polytech network. The network is composed of 15 engineering schools/universities. The document also provides information about the supervisor. Please contact the PhD supervisor by email for further information regarding your application.

Supervisor information	
Family name	BENABDESLEM
First name	Khalid
Email	Khalid.benabdeslem@univ-lyon1.fr
Web reference	http://perso.univ-lyon1.fr/khalid.benabdeslem
Lab name	LIRIS
Lab web site	https://liris.cnrs.fr
Polytech name	POLYTECH LYON
University name	University LYON 1
Country	France

PhD information	
Title	Multi-view Learning for Inherited Cardiomyopathy Analysis.
Main topics regards to CSC list (3 topics at maximum)	I-12: Understanding models and intelligent systems III-4: Explosion of serious illnesses and prevention

Required skills in science and engineering	Computer sciences, Artificial intelligence, Statistics, Health.

Subject description (two pages maximum including biblio)

1. Context

The approach through the thesis project is part of a global vision to analyze different complex Inherited cardiac conditions (ICC) are severe and misunderstood diseases that affect millions of young patients worldwide. More than a thousand mutations have been identified in many genes, indicating that diverse molecules and pathways cause channelopathies and hypertrophic, dilated, restrictive, and arrhythmogenic cardiomyopathies. These mutations are often patient specific, relevant for a person or family. They may also be group specific, or the hallmark of sub-specific populations of patients. However, genetic heterogeneity and allelic variation together with variation in clinical expression, pose substantial challenges for clinicians faced with their patients.

The thesis proposes a diagnostic support intelligent system to optimize diagnosis of ICC. The proposal aid tool will structure the clinical decision for patients with suspected ICC, precisely define the question to be addressed, identify exhaustive actions that may be undertaken and their potentials outcomes, estimates the probability that events will occur and their associated costs and values.

2. Objectives

The objective of this thesis is to analyze different complex data (ECG forms, genetic variants and other clinical information…) with data mining and machine learning techniques for decision aid in the broad sense. The ECG signals and the clinical and biological data will be restructured in the form of matrix representations, which may contain all kinds of relevant information or not. Among these information, two sets of dimensions are of primary interest for clinicians: Cardiac behaviors, via ECG signals, and genetic factors, via DNA records of patients. These dimensions are very important indicators that provide synoptic information on the overall performance of a medical diagnosis of a patient or more generally a group of patients.

Given the complex nature of the data collected on patients, the needs expressed by the clinicians cover the following aspects, which are the operational and scientific objectives of this work:

- Detection of cardiac anomalies from ECG signals.
- Extraction of patient profiles from complex data (ECG, genetic and clinical data).
- Automatic diagnosis from complex data, with good precision
- Need of therapeutic decision support based on diagnosis provided by AI solutions.
- Need of interpretation process by selecting genes that explain the diagnosis of different pathologies.
- Need of scalable machine learning tools for dealing with huge amounts of high dimensional data.

3. Scientific program

The operational objective of the scientific program behind this thesis is to demonstrate, with mathematical and computer tools, the interaction between the clinic and genetics, which represents the main challenge of the project. This demonstration will involve close collaborations and fruitful exchanges between researchers from the DM2L (Data Mining and Machine learning) team of LIRIS Lab and cardiologists and geneticists of HCL (Hospital of Lyon). The aim is to develop the best relationships between the fundamental aspect, by new machine learning algorithms and the decision-making aspect that seriously concerns clinicians.

As part of this thesis, we propose to develop data mining tools based on statistical learning models for the automatic diagnosis of pathologies of inherited cardiomyopathy. The main goal is to help with the therapeutic decision of these serious, complex and poorly understood diseases actually. To do this, we adopt a data-driven approach that will be used to build our models, by four main tasks:

3.1. Cardiac anomaly detection from ECG data: In this task, we propose to develop a global approach which allows to detect ECG data, showing cardiac anomalies and then, model this particular form according to the characteristics of the signals.

3.2. Profiling of patients from complex data: The objective is to combine, simultaneously, EGC signals and biological characteristics to extract homogeneous clusters of patients that can provide a complete picture expressing similar symptoms, for making good decisions. We propose to carry out this task on two steps : coding the genetic variants and integrating them, with other types of data, on an unsupervised learning approach in order to extract accurate profiles of patients.

3.3. Diagnosis of inherited cardiomyopathy and channelopathy: As reported in task 3.2, the data are complex. In fact, each patient is described by several ECG leads, several genes and several clinical data (sex, age, ...). In addition, the classes about ICC are multiple (dilated, hypertrophic, restrictive, non-compaction, long QT and Brugada syndromes, catecholaminergic polymorphic ventricular tachycardia, idiopathic ventricular fibrillation). To treat such heterogeneous data to make a diagnosis, is part of a special case of supervised machine learning, which is **multi-view learning**.

3.4. Factor selection for explaining the diagnosis of inherited cardiomyopathy: This task is important for explaining the diagnosis and can be carried out in machine learning by feature selection methods. The objective is to determine the relevant factors from ECG, genetic and clinical spaces, allowing to explain a given diagnosis.

4. References

- S-E. Benkabou, **K. Benabdeslem**, V. Kraus, K. Bourhis, B. Canitia. *Local Anomaly Detection for Multivariate Time Series by Temporal Dependency Based on Poisson Model*, **IEEE Transactions on Neural Networks and Learning Systems,** 33(11), 6701-6711, 2022.

- K. Benabdeslem, D-E-K. Mansouri and R. Makkhongkaew. *sCOs: Semi-supervised co-selection by a similarity preserving approach*, **IEEE Transactions on Knowledge and Data Engineering,** 34(6): 2899-2911, 2022.

- T. Ranvier, **K. Benabdeslem**, K. Bourhis, B. Canitia. *Deep Multi-View Learning for Tire Recommendation.* In the proceeding of **IEEE IJCNN**: International Joint Conference on Neural Network, 18-22 July 2021, Shenzhen (China), (Online).