



Avis de Soutenance

Madame Carine BOU RJEILY

Présentera ses travaux en soutenance

Soutenance prévue le **vendredi 04 octobre 2019** à 13h30
Lieu : UTBM Site de Belfort Rue Thierry Mieg 90010 BELFORT Cedex
Salle : I102

Titre des travaux : Fouille de données et apprentissage pour l'extraction de marqueurs pour améliorer le raisonnement des plateformes de suivi médical

Ecole doctorale : SPIM - Sciences Physiques pour l'Ingénieur et Microtechniques

Section CNU : 27

Unité de recherche : Laboratoire de Nanomédecine, Imagerie, Thérapeuthique

Directeur de thèse : Amir HAJJAM EL HASSANI

Codirecteur de thèse : Emmanuel ANDRES HDR NON HDR

Soutenance : Publique A huis clos

Membres du jury :

<u>Nom</u>	<u>Qualité</u>	<u>Etablissement</u>	<u>Rôle</u>
M. Amir HAJJAM EL HASSANI	Maître de Conférences	Université Bourgogne - Franche-Comté	Directeur de thèse
M. Emmanuel ANDRES	Professeur	Université de Strasbourg	Co-directeur de thèse
M. Georges BADR	Maître de Conférences	Université Antonine	Co-directeur de thèse
Mme Florence SEDES	Professeur	IRIT - Université Paul Sabatier	Examineur
Mme Cécilia ZANNI-MERK	Professeur	LITIS / Equipe MIND INSA ROUEN	Rapporteur
Mme Marie-Hélène ABEL	Professeur	Heudiasyc UTC, Université de Technologie de Compiègne	Rapporteur

Résumé de la thèse (en français) :

Selon l'Organisation mondiale de la santé, environ 31% des décès dans le monde sont causés par des maladies cardiaques chaque année. L'exploration de données est un processus d'extraction intéressant d'informations non triviales, implicites et potentiellement utiles, à partir de grands ensembles de données. L'exploration de données médicales est la science qui consiste à examiner des données médicales (signes vitaux) pour explorer des informations importantes. L'analyse et l'interprétation des données complexes dans un diagnostic thérapeutique approprié avec les bons résultats, est une tâche assez ardue. Néanmoins, le fait qu'il soit possible de combiner ces facteurs jusqu'à un certain point et d'extraire un plan de traitement, de prévention et de rétablissement généralement couronné de succès, est un signe des avantages à venir. Grâce à cela, il est maintenant possible d'améliorer la qualité de vie des patients, de prévenir une aggravation de la maladie tout en maintenant les coûts médicaux à la baisse. Cela explique la popularité croissante de l'utilisation et de l'application des techniques d'apprentissage automatique pour analyser, prédire et classer les données médicales. Dans une première contribution, nous avons étudié de nombreux algorithmes de motifs séquentiels qui sont des techniques prometteuses pour l'exploration de données. Nous les avons classés afin de choisir un algorithme approprié pour prédire les classes d'insuffisance cardiaque et sa présence. Après avoir comparé tous les algorithmes et les avoir mis en œuvre sur le même ensemble de données médicales, le CPT+, un algorithme de prédiction de séquence, a été choisi en donnant les résultats les plus précis avec une précision de 90,5% dans la prédiction de l'insuffisance cardiaque et de ses classes. Après, nous avons basculé nos études vers une stratégie de séries chronologiques et nous avons utilisé des données réelles extraites de patients réels. 5 paramètres ont été extraits de 3 patients au cours de quelques années. L'algorithme Random Tree a donné plus de 85% de prédictions correctes de l'insuffisance cardiaque 7 jours à l'avance.

Abstract (in English):

The World Health Organization accords that about 31 % of deaths worldwide are caused by heart diseases every year. Data mining is a process of extracting interesting non-trivial, previously unknown and potentially useful information from huge amount of data. Medical data mining is the science of investigating medical data (i.e. vital signs) to explore significant information. Analyzing and interpreting the huge amount of complicated data into an appropriate therapeutic diagnosis with the right results is quite challenging task. Still, the fact that it is possible to combine these factors up to a certain point and extract a usually successful treatment, prevention and recovery plan is a sign of the good things to come. Thanks to that, it is now possible to improve patients' quality of life, prevent condition worsening while maintaining medical costs at the decrease. This explains the increasing popularity in the usage and application of machine learning techniques to analyze, predict and classify medical data. As a first contribution, we studied many sequential patterns algorithms that are promising techniques in exploring data and we classified them in order to choose an appropriate one for predicting Heart Failure classes and presence. After comparing all the algorithms and implementing them on the same medical dataset, the CPT+ a sequence prediction algorithm has been chosen as it gave the most accurate results reaching an accuracy of 90.5% in predicting heart failure and its classes. Thereafter, we switched our studies to time series strategy, and worked on real data extracted from real patients. 5 parameters were extracted from 3 patients over the course of a few years. The Random Tree algorithm yielded more the 85% correct predictions of heart failure 7 days prior.