# Conclusion et perspective

Ce mémoire s’est intéressée au problème de classification mono- et multi-label dans un contexte médical lié au suivi des maladies cardio-vasculaires (MCV). L’objectif de cette thèse était de proposer des approches intelligentes qui anticipent l’évolution des MCV chez des patients dans leur lieu de vie quotidien sans imposer des contraintes gênantes. Ces approches se basent sur des mesures de signaux physiologiques en temps réel pour les analyser et en sortir une décision. Dans ce mémoire, nous avons opté pour l’utilisation de deux formalismes différents pour répondre à ce problème. Le premier adopte la notion de classification mono-label qui vise la prédiction d’épisode de fibrillation atriale (FA) basée sur la théorie de fonctions de croyance, et le second adopte la notion multi-label pour pouvoir effectuer un diagnostic multi-pathologique. Le défi scientifique à relever dans ce mémoire était la proposition d’approches de classification permettant de répondre aux contraintes de données manquantes par moment, hétérogènes, imprécises et indirectement liées au phénomène observé.

Le manuscrit commence par une introduction des approches technologique dont l’intelligence artificielle, le système de santé, le Bigdata ainsi que le chatbot et leurs capital intérêt applique aux MCV démontrant l’ampleur de leur impact sur la santé et mettant en évidence les enjeux des travaux de recherche tout en introduisant le matériel utilisé dans ces travaux qui demeure principalement la base de données médicales MIMIC III. Ensuite, Le manuscrit introduit la théorie des fonctions de croyance, étant l’outil technique de base de la majorité des travaux développés. Quatre contributions sont présentées par la suite. La première porte sur une approche de classification visant à prédire un épisode de la FA. Cette méthode est basée sur la théorie des fonctions de croyance, qui est une théorie permettant de modéliser l’information d’une manière différente de la théorie de probabilités en tenant en compte des spécificités de l’application relatives au manque d’information.

La mise en œuvre de la méthode développée a donné, après optimisation des paramètres et sur un ensemble de patients test et un groupe de contrôle, une précision de 79.3%, une sensibilité de 74.2%, une spécificité de 83.7% ainsi qu’une portée de prédiction d’environ 2 heures et 10 minutes avant l’occurrence de l’épisode. On a apporté par la suite des développements améliorant les performances de cette méthode en proposant une nouvelle modélisation de l’information et une option de rejet supplémentaire. Le manuscrit introduit aussi une approche originale pour la fusion des données se traduisant par une méthode de combinaison plus flexible que la règle de YAGER. On a montré qu’on peut atteindre avec ces améliorations des précisions allant jusqu’à 95.8% pour un taux de rejet d’environ 38.21%. La méthode est évaluée sur la base de données MIMIC III.

Par ailleurs, on a introduit la notion de classification multi-label avant de l’appliquer au diagnostic multi-pathologique. Les signaux physiologiques de la base de données médicale MIMIC III ont été utilisés pour faire le diagnostic des éventuelles complications cardiaques relatives à 55 maladies cardio-vasculaires. On a aussi introduit une approche multi-label basée sur la théorie des fonctions de croyance qu’on a validée grâce aux données extraites de la base médicale MIMIC III.

La méthode a été comparée à des méthodes multi-label classiques basées sur une transformation de plusieurs méthodes de classification mono-label. Il s’est avéré que la méthode proposée est meilleure que toutes les autres méthodes en utilisant plusieurs critères d’évaluation. Ceci montre la capacité de la méthode proposée à traiter les données manquantes. Enfin, on a proposé une méthode originale appelée COLEX pour la classification multi-label puis on l’a appliquée à notre problème de diagnostic. Cette méthode s’inscrit dans le cadre des méthodes de transformation, comme elle transforme le problème de classification multi-label en plusieurs sous-problèmes mono-label. Elle est une généralisation de la méthode RAkEL qui se base sur l’optimisation du choix des *k*-labelsets en considérant les corrélations existantes entre les labels. En effet, la méthode permet un réglage automatique de la répartition des *k*-labelsets sur les *n* modèles de la méthode RAkEL. Les résultats obtenus sur des données du monde réel ont montrées l’efficacité de notre méthode par rapport à la méthode RAkEL classique et à d’autres méthodes de la littérature.

Des perspectives sont envisageables pour la suite des travaux de recherche sur ce sujet :

* D’abord, les approches proposées extraient des caractéristiques statistiques simples des signaux physiologiques. Dans des travaux futurs, il est envisageable d’extraire des caractéristiques en utilisant d’autres approches linéaires ou non linéaires, telles que celles basées sur l’analyse par composantes principales.
* Les caractéristiques extraites de la base d’apprentissage sont modélisées individuellement par des distributions de probabilité normales. Il serait intéressant de proposer des distributions de probabilité plus adéquates, adaptées à chaque caractéristique. On peut par exemple se servir de l’estimation à noyau des fonctions de densité ou de concevoir une sélection automatique parmi plusieurs distributions de probabilité standards bien connues pour chaque caractéristique indépendamment. Il est intéressant aussi d’envisager une modélisation multidimensionnelle par signal, en considérant toutes les caractéristiques d’un même signal comme un vecteur. Cela permettra de profiter de la dépendance entre les caractéristiques et préservera la robustesse des méthodes aux signaux manquants.
* Une autre perspective consiste à adapter l’approche de la théorie des fonctions de croyance à la classification multi-classe permettant ainsi de résoudre des problèmes de classification multi-label en appliquant une méthode de transformation en plusieurs problèmes de classification multi-classe. Cette extension doit prendre en considération la conception d’une modélisation adéquate pour chaque sous-ensemble de classes.
* D’autre part, la théorie des fonctions de croyance permet d’attribuer des poids à des sous-ensembles de l’ensemble des labels possibles. Cela laisse présager que cette théorie pourrait être convenable à la conception d’une méthode de classification multi-label par adaptation, agissant directement sur les données multi-label.
* Il est aussi possible d’intégrer dans le modèle de nouvelles sources d’information qui s’appuient sur des données statistiques de la maladie (prévalence, incidence, symptômes connus, etc.) ou catégoriques propres au patient (séjours précédents, fiche de soins, antécédents médicaux, etc.).
* De nos jours, grâce à la domotique et l’affluence des objets connectés dans la vie quotidienne, il est possible de collecter de grands volumes de données liées d’une manière directe ou indirecte à l’état de santé d’un sujet. En présence de ces données, il est intéressant d’utiliser l’apprentissage profond (deep Learning) à plusieurs niveaux comme au niveau de l’extraction de caractéristiques, le remplissage des données manquantes ou la classification elle-même. En fonction du volume de données, il est aussi possible que plusieurs de ces tâches soient accomplies simultanément par un réseau de neurones utilisant l’apprentissage profond. Il serait alors intéressant d’étudier la transformation des signaux en images cohérentes pour tirer profit de ces réseaux.
* Les approches proposées permettent le suivi d’un patient quelconque grâce à un classifier pré-entraîné. Il est judicieux de calibrer le système aux informations du patient, en exploitant le transfert d’apprentissage.
* L’apprentissage proposé utilise la base de données MIMIC III, formée de signaux collectés en unités de soins intensifs. Cela nous a permis de valider nos approches grâce à des signaux réels. Toutefois, ces signaux ne sont pas représentatifs du cadre applicatif auquel est dédié le système, à savoir la surveillance à domicile. Une collaboration avec une équipe de cardiologues est envisagée pour l’accès à des données plus réalistes.
* Enfin, les développements menés à ce jour se situent au niveau méthodologique. D’un point de vue technologique, d’autres perspectives vise à développer un système de surveillance à domicile respectant les contraintes pratiques d’acceptabilité, éthiques et économiques. Il est nécessaire que le produit soit accepté et validé par la communauté visée.