

## DIPLÔME NATIONAL DE DOCTORAT

(Arrêté du 25 mai 2016)

Date de la soutenance : **20 décembre 2024**

Nom de famille et prénom de l'auteur.e : **Monsieur Abdelhamid GADDARI**

Titre de la thèse : Analyse et Prédiction des Parcours Patients dans le contexte des Complémentaires Santé

### Résumé



Ce travail de thèse s'inscrit dans la catégorie de la recherche en informatique de santé, en particulier l'analyse et la prédiction des parcours patients, qui sont les séquences des actes médicaux consommés par les patients au fil du temps. Notre objectif est de proposer une approche innovante pour l'exploitation des données de parcours de soins afin de réaliser non seulement une classification binaire, mais aussi multi-label. Nous concevons également une nouvelle approche de vectorisation et représentation sémantique exclusivement pour le domaine médical français, qui permettra d'exploiter un autre aspect des parcours patients afin d'améliorer la performance prédictive de notre approche proposée. Notre recherche s'inscrit dans le cadre des travaux de CEGEDIM ASSURANCES, une business unit du groupe CEGEDIM qui fournit des logiciels et des services pour les secteurs de l'assurance maladie complémentaire et de la gestion des risques en France. En analysant le parcours de soins et en utilisant l'approche que nous proposons, nous pouvons extraire des informations précieuses et identifier des patterns dans les parcours médicaux des patients afin de prédire des événements médicaux potentiels ou la consommation médicale à venir. Cela permettra aux assureurs de prévoir les futures demandes de soins de santé et donc de négocier de meilleurs tarifs avec les prestataires de soins de santé, ce qui permettra une planification financière précise, des modèles de tarification équitables et une réduction des coûts. En outre, ça permettra aux assureurs privés de concevoir des plans de santé personnalisés qui répondent aux besoins spécifiques des patients, en veillant à ce qu'ils reçoivent les soins adéquats au bon moment afin de prévenir la progression de la maladie. Enfin, l'offre de programmes de soins préventifs et de produits et services de santé personnalisés renforce les relations avec les clients, améliore leur satisfaction et réduit l'attrition. Dans ce travail, nous visons à développer une approche permettant d'analyser les parcours patients et de prédire les événements médicaux ou les traitements à venir, sur la base d'un large portefeuille de remboursements. Pour atteindre cet objectif, nous proposons tout d'abord un

nouveau modèle basé sur les LSTM qui tient compte de la notion temporelle et qui permet de réaliser de la classification binaire et multi-label. Le modèle proposé est ensuite étendu par un autre aspect des parcours de soins, à savoir des informations supplémentaires provenant d'un clustering flou du même portefeuille. Nous démontrons que l'approche proposée est plus performante que les méthodes traditionnelles et d'apprentissage profond dans la prédiction médicale binaire et multi-label. Par la suite, nous améliorons la performance prédictive de l'approche proposée en exploitant un aspect supplémentaire des parcours patients, qui consiste en une description textuelle détaillée des traitements médicaux consommés. Ceci est réalisé grâce à la conception de F-BERTMed, une nouvelle approche de vectorisation et de représentation sémantique de phrases pour le domaine médical français. Celle-ci présente des avantages significatifs par rapport aux méthodes de l'état de l'art du traitement automatique du langage naturel (TAL). F-BERTMed est basé sur FlauBERT, dont le pré-entraînement utilisant la tâche MLM (Modélisation Masqué du Langage) a été étendu sur des textes médicaux français avant d'être fine-tuné sur les tâches NLI (Inférence du Langage Naturel) et STS (Similarité Sémantique Textuelle). Nous démontrons enfin que l'utilisation de F-BERTMed pour générer une nouvelle représentation des parcours patients améliore les performances prédictives de notre modèle proposé pour les tâches de classification binaire et multi-label.

**Mots-clés :** Parcours Patients, Prédiction Médicale, Assurance Complémentaire de Santé, LSTM Temporel, TAL, Représentation Sémantique