

# ICDiscovery : Aide à l’annotation par une méthode de budget pour le codage CIM-9 des textes hospitaliers

Leonardo Moros<sup>\*,\*\*\*</sup>, Jérôme Azé<sup>\*</sup>, Sandra Bringay<sup>\*,\*\*</sup>  
Pascal Poncelet<sup>\*</sup>, Maximilien Servajean<sup>\*,\*\*</sup>, Caroline Dunoyer<sup>\*\*\*,\*\*\*\*</sup>

\* LIRMM UMR 5506, Université de Montpellier, CNRS, Montpellier, France  
prenom.nom@lirmm.fr

\*\* Groupe AMIS, Université Paul Valéry, Montpellier, France  
prenom.nom@univ-montp3.fr

\*\*\* Département d’Information Médicale, CHU Montpellier, Montpellier, France  
prenom.nom@chu-montpellier.fr

\*\*\*\* IDESP, UMR UA11, INSERM - Université de Montpellier, Montpellier, France  
prenom.nom@umontpellier.fr

**Résumé.** Le codage médical est une tâche liée à la facturation clinique, visant à annoter des textes, généralement non structurés, avec des codes décrivant les diagnostics et les traitements d’un patient. Cette tâche réalisée par des spécialistes du codage, est connue pour être très difficile, en raison de la grande quantité de codes et de la longueur des documents. ICDiscovery est un outil d’aide au codage basé sur une approche d’apprentissage automatique par budget, qui propose un nombre de codes différents à associer à des documents selon plusieurs stratégies et qui explique les prédictions en identifiant les parties de textes qui ont permis la prédiction des codes.

## 1 Introduction

Les professionnels de santé documentent minutieusement chaque rencontre avec les patients dans leurs dossiers médicaux. Ils produisent de nombreux documents structurés et semi-structurés contenant des informations sur les traitements, les procédures et les diagnostics effectués. Afin d’obtenir des financements, les établissements de santé doivent associer aux séjours de patients des codes de facturation, issus de la Classification Internationale des Maladies (CIM). Ce codage, initialement réalisé à des fins médico-économiques peut avoir d’autres finalités que la facturation, telles que l’amélioration de la prise en charge du patient, la prédiction de l’évolution des soins, etc. Actuellement, cette tâche est effectuée manuellement par des spécialistes du codage, c’est une activité très complexe, fastidieuse, subjective, coûteuse, chronophage et sujette à erreurs.

De nombreux travaux ont proposé des systèmes automatiques pour cette activité de codage. Ces dernières années, les approches utilisant des modèles d’apprentissage profond ont obtenu les meilleurs résultats. Les réseaux neuronaux convolutifs (CNN) et récurrents (RNN) avec des mécanismes d’attention (Xie et Xing, 2018; Mullenbach et al., 2018; Vu et al., 2020) correspondent à l’état de l’art actuel.