

# Segmentation automatique des rapports médicaux en utilisant les réseaux de neurones convolutionnels

Walid Zeghdaoui\*,\*\* Frederik Joly\*\*  
Omar Boussaid\* Fadila Bentayeb\*

\*Université de Lyon, Université Lyon 2, ERIC EA 3083  
5, Av. Pierre Mendès-France, 69676 Bron, France

{walid.zeghdaoui, omar.boussaid, fadila.bentayeb}@univ-lyon2.fr

\*\*Sword Group - 9, Av. Charles de Gaulle, 69370 St-Didier-au-Mont-d'Or, France  
{walid.zeghdaoui, frederik.joly}@sword-group.com

**Résumé.** L'un des défis majeurs de la médecine de précision est d'orienter la recherche et le développement de solutions thérapeutiques spécifiques en exploitant les informations contenues dans les rapports médicaux. Ces rapports sont souvent maintenus sous forme de textes non structurés et constituent un volume important de données. L'extraction de connaissances à l'aide des techniques de traitement automatique de langage naturel (TALN) pourrait donc aider à améliorer les soins de santé et la prise de décisions médicales. Cette étape est souvent précédée par une phase de segmentation de textes pour bien identifier les parties du rapport d'un grand intérêt. Dans cet article, nous présentons notre système automatique de segmentation de rapports médicaux en sections sémantiques et qui consiste en deux étapes complémentaires. D'abord un algorithme basé sur la détection de titres pour l'identification de certaines sections. La deuxième partie est une tâche de classification de phrases supervisée basée sur des algorithmes d'apprentissage profond. Ce système a été testé sur 500 rapports et a atteint une précision de classification de plus de 96%.

## 1 Introduction

En dépit des efforts consacrés pour la saisie de données cliniques dans un format structuré, la plupart des rapports médicaux sont aujourd'hui maintenus en textes libres. La quantité de ces données a augmenté considérablement au cours de la dernière décennie et risque de se multiplier dans les années à venir. Ces données contiennent une mine d'informations et représentent une riche base de connaissances qui doit être impérativement exploitée pour améliorer le processus de soins de santé. En effet, l'utilisation de ces informations permet d'aider les médecins sur plusieurs plans, notamment faciliter la prise de décisions médicales. Par exemple l'identification du diagnostic le plus approprié pour un patient, mais aussi pour faire progresser la recherche clinique, car la communauté médicale s'efforce constamment de trouver de nouveaux moyens pour mener des recherches qui s'inscrivent dans la lutte contre les maladies. Afin de faciliter leur tâche, notre travail consiste à rendre ces informations facilement accessibles. Ainsi l'utilisation des techniques de recherche et d'extraction d'information est devenue