# Exploitation des dépendances entre labels pour la classification de textes multi-labels par le biais de transformeurs

Haytame Fallah[*,***], Patrice Bellot[*], Elisabeth Murisasco[**], Emmanuel Bruno[**]

[*] Aix-Marseille Université, Université de Toulon, CNRS, LIS, Marseille, France
[**] Université de Toulon, Aix-Marseille Université, CNRS, LIS, Toulon, France
[***] Hyperbios, Aix-en-Provence, France

**Résumé.** Nous présentons une nouvelle approche pour améliorer et adapter les transformeurs pour la classification multi-labels de textes. Les dépendances entre les labels sont un facteur important dans le contexte multi-labels. Les stratégies que nous proposons tirent profit des co-occurrences entre labels. Notre première approche consiste à mettre à jour l'activation de chaque label par une somme pondérée de toutes les activations par les probabilités d'occurrence. La deuxième méthode proposée consiste à inclure les activations de tous les labels dans la prédiction, en utilisant une approche similaire au mécanisme de 'self-attention'. Les jeux de données multi-labels les plus connus ont tendance à avoir une faible cardinalité, nous proposons un nouveau jeu de données, appelé 'arXiv-ACM', composé de résumés scientifiques d'arXiv, étiquetés avec leurs mots-clés ACM. Nous montrons que nos approches contribuent à un gain de performance, établissant un nouvel état de l'art pour les jeux de données étudiés.

## 1 Introduction

La classification multi-labels peut être considérée comme une généralisation de la classification multi-classes. Dans la classification multi-labels de texte *(CMLT)*, l'objectif est d'associer un ou plusieurs labels au texte d'entrée. C'est une tâche importante pour différentes applications telles que la réponse aux questions (Wu et al., 2019; Sahu et al., 2019) où les questions peuvent contenir plus d'un sujet, ou la reconnaissance d'entités (Remolona et al., 2017; de Souza et al., 2020) où les entités peuvent avoir plusieurs catégories sémantiques. Dans ce contexte, le nombre de labels à prédire est plutôt limité, ce qui diffère de la CMLT extrême (Liu et al., 2017; Yu et al., 2019; Shen et al., 2020). Plusieurs méthodes ont été proposées pour la CMLT et peuvent être divisées en deux familles : les méthodes de transformation de problèmes et les méthodes d'ensemble. Les méthodes de transformation de problèmes (Tsoumakas et al., 2010; Luaces et al., 2012) visent à "transformer" le jeu de données pour changer le problème en une classification multi-classes à label unique. Dans les méthodes d'ensemble (Tsoumakas et Vlahavas, 2007; Saini et Ghosh, 2017), plusieurs classifieurs sont formés pour prédire la présence d'un label, puis combinés pour capturer tous les labels présents. En transformant le