

## ***ABSTRACT***

### **Generalization of LR-GCN For Multi-label Text Classification**

by

Dzaki Dzikril Hakim 20/454530/PA/19561

The rapid growth of multilingual digital text has intensified the need for Multi-Label Text Classification (MLTC) models that generalize across domains, label structures, and languages. This research investigates the generalization ability of the Label-Representative Graph Convolutional Network (LR-GCN), a model that incorporates label co-occurrence graphs with label and document embeddings. This research uses four diverse and challenging datasets, three Chinese CAIL2019 Feature Recognition subsets and the English LexGLUE EUR-Lex dataset. These differ to previously tested datasets in LR-GCN with its legal domain specificity and language. To account for the multilingual nature of the some of the data, the original English transformer encoders were replaced with Chinese equivalents. The findings show that LR-GCN is consistently competitive with a recent MLTC model, achieving lower Hamming Loss across all Chinese datasets. It also outperforms LSAN, another label-focused MLTC model, on all P@k and nDCG@k metrics in LexGLUE EUR-Lex. These results show that LR-GCN not only maintains competitive predictive accuracy across languages and domains but also benefits from its label graph structure, especially in settings with sparse or uneven label co-occurrence patterns.

**Keywords:** *Multi-Label Text Classification, Graph Convolutional Network, Label Co-occurrence, Transformer Embeddings*