

# TMUNLP at the NTCIR-17 MedNLP-SC Task

Yong-Zhen Huang<sup>1,2</sup>, Eugene Sy<sup>1</sup>, Yi-Xuan Lin<sup>1</sup>, Yu-Lun Hsieh<sup>1</sup>, Yung-Chun Chang<sup>1,\*</sup>

1. Graduate Institute of Data Science, Taipei Medical University, Taiwan

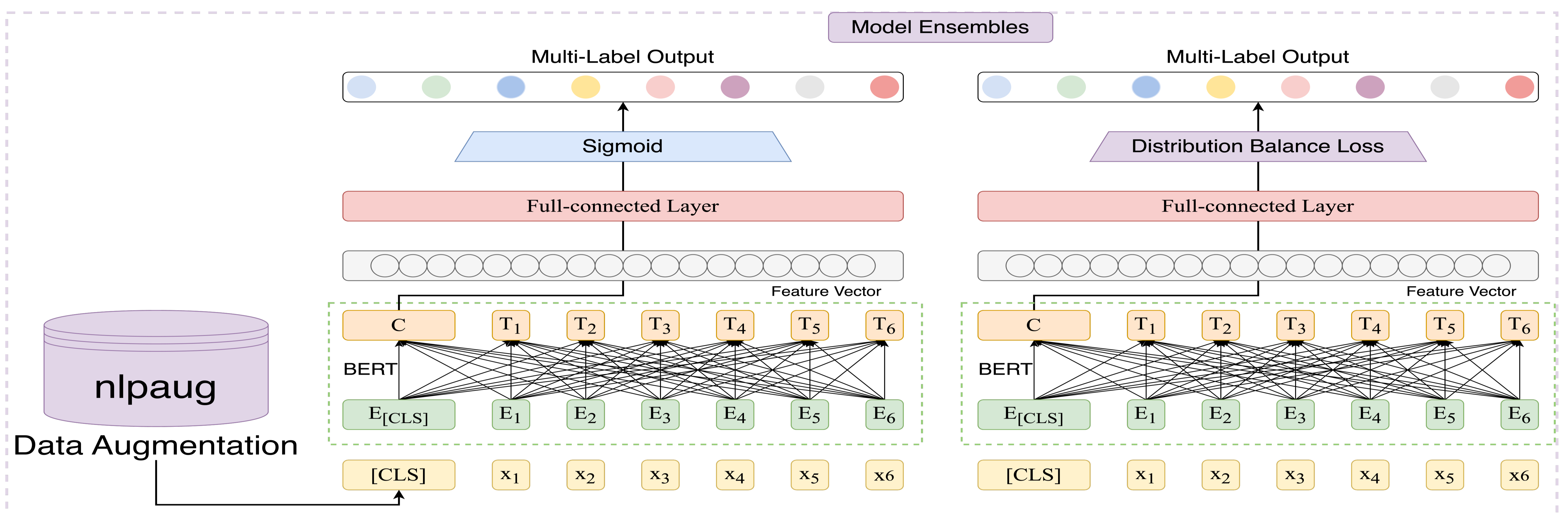
2. Department of Nursing, National Taiwan University Cancer

## 【Introduction】

In the NTCIR-17 MedNLP-SC, we focused on the critical challenge of detecting adverse drug events (ADE) from social media, using advanced natural language processing techniques. Our goal was to develop systems capable of accurately identifying and classifying ADE information, including a wide range of symptoms, from English-language social media posts. This task addresses the need for practical systems to support medical services, with a particular emphasis on the complexities of multi-label classification in the context of ADE detection.

## 【Methods】

We utilized BERT-base-uncased and ClinicalDistilBERT for their adeptness in processing medical texts. To tackle label co-occurrence issues, we implemented Distribution Balance Loss, enhancing focus on complex and underrepresented cases. Additionally, we used 'nlpaug' for synonym-based data augmentation, targeting minority categories. Finally, model ensembles, selected through k-fold cross-validation, were used for refined and accurate predictions.



## 【Result】

Our submissions excel in the SM-ADE-EN test set, particularly distinguished by their per ADE label F1-scores, which are essential for precise ADE detection. Submission-1 stands out with the highest per ADE label F1-score of 0.71, underscoring its adeptness at identifying diverse ADE symptoms. Despite XLM-RALL's slight edge in exact match accuracy, our models demonstrate enhanced monolingual efficiency and achieve state-of-the-art results through our integration of ensemble techniques, distribution balance loss, and data augmentation.

Systems	Binary Score		(Full) Per Label Scores		(Individual) Per Class Scores	Exact Match Accuracy Score
	Class	F1	Class	F1	F1 (macro avg.)	
BERT	ADE	0.71	0	0.99	0.41	0.79
	non-ADE	0.89	1	0.66		
XLM-R	ADE	0.63	0	0.99	0.26	0.76
	non-ADE	0.88	1	0.57		
XLM-RALL	ADE	0.78	0	0.99	0.61	0.83
	non-ADE	0.91	1	0.76		
Submission-1	ADE	0.79	0	0.99	0.71	0.83
	non-ADE	0.90	1	0.76		
Submission-2	ADE	0.78	0	0.99	0.70	0.82
	non-ADE	0.90	1	0.76		
Submission-3	ADE	0.79	0	0.99	0.69	0.82
	non-ADE	0.90	1	0.76		

## 【Conclusion】

Our study improved adverse drug event prediction in multi-label classification by using Distribution Balance Loss and Data Augmentation, enhancing Macro F1-scores for minority labels. Despite these gains, the exact accuracy impact was modest, indicating potential areas for future improvement in balancing label performance. Further analysis is needed to optimize ensemble model contributions for more equitable classification outcomes.

## 【Acknowledgements】

This study was supported by the National Science and Technology Council under grant 112-2410-H-038-007- and 112-2622-E-038-001-, as well as the National Health Research Institutes NHRI-12A1-PHCO-1823244.