



# NTTMUNSW System for n2c2 Track1: Cohort Selection for Clinical Trails



Feng-Duo Wang<sup>1</sup>, Chih-Wei Chen<sup>2</sup>, Hong-Jie Dai<sup>1</sup>, Chu-Hsien Su<sup>3</sup>, Chi-Shin Wu<sup>3</sup>, Jitendra Jonnagaddala<sup>4</sup>

<sup>1</sup>Department of Computer Science and Information Engineering, National Taitung University, Taitung, Taiwan R.O.C.,

<sup>2</sup>Graduate Institute of Biomedical Informatics, Taipei Medical University, Taipei, Taiwan R.O.C.,

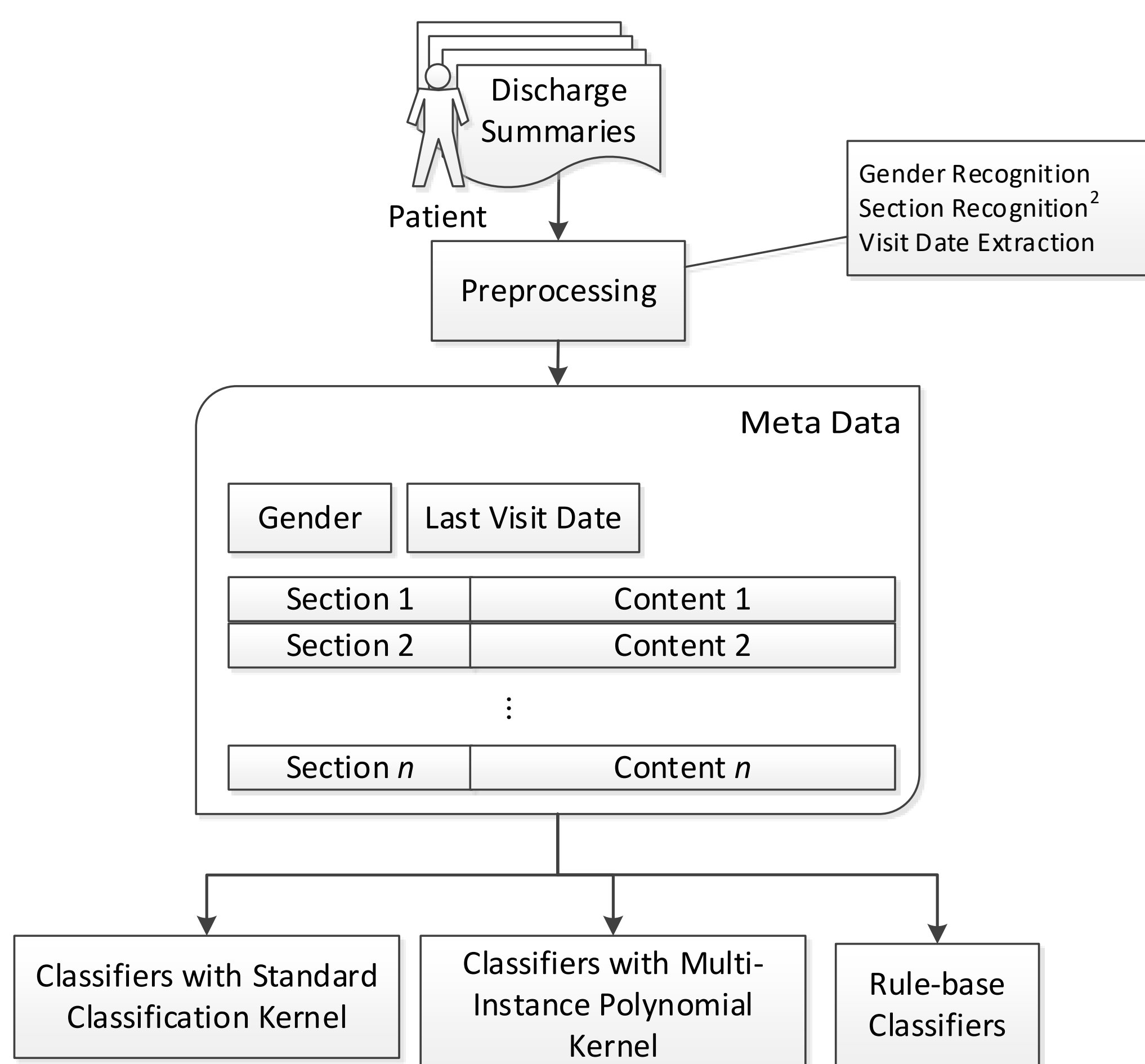
<sup>3</sup>Department of Psychiatry, National Taiwan University Hospital, Taipei, Taiwan R.O.C.,

<sup>4</sup>School of Public Health and Community Medicine, UNSW Sydney, Australia

## Introduction

As part of the 2018 n2c2 shared task, NTTMUNSW team developed a hybrid system combined rule-based methods with machine learning based approaches to automatically identify potential patients for clinical trials to aid the process of manually chart review of electronic health records (EHRs).

## Methods



We formulated the problem as a **multiple instance (MI)** classification problem<sup>1</sup>. The set of EHRs from each patient is considered as a labeled bag, each containing at most five longitudinal records. The rationale is that if a bag is labeled with “not met”, all the EHRs in a labelled bag should be “not met” as well. On the other hand, a bag is labeled “met” if there is at least one record in it which meets the criterion. For each selection criterion, we followed the formulation to develop classifiers.

## Results

Configuration	Micro-F	AUC
Run 1 (Multi-instance Learning Hybrid)	<b>0.8765</b>	<b>0.8807</b>
Run 2 (Rule-based only)	0.8641	0.8666
Run 3 (Standard classification method only)	0.8469	0.8527

The MI learning configuration achieved the best overall micro-F-score and AUC. It also had best precision and recall on the “met” criterion and the recall on the “not met” criterion. The standard classification method based on the aggregated-based machine learning models had the lowest overall precision/recall/f-score and specificity on both criteria.

## Conclusion

The proposed MI method is a better approach than traditional machine learning models for clinical trial pre-screening of patients with a set of EHRs.

## References

1. Dietterich, T.G., *et al.*, *Solving the multiple instance problem with axis-parallel rectangles*. Artificial intelligence, 1997.
2. Chang, N.-W., *et al.*, *A context-aware approach for progression tracking of medical concepts in electronic medical records*. J. Biomedical Informatics, 2015.

## Acknowledgements

This work was supported by the Ministry of Science and Technology of Taiwan [grant numbers MOST-105-2221-E-143-003 and MOST-106-2221-E-143-007-MY3].