

## 11100

## Identification of ICD-code misclassifications in cardiac disease using natural language processing

Falter M Doctor, Godderis D Mr, Scherrenberg M Doctor, Kizilkilic SE Doctor, Xu L Miss, Tukanov E Mr, Neven F Professor, Dendale P Professor

Hasselt University, Hasselt, Belgium

**Funding Acknowledgements:** Type of funding sources: Public grant(s) – National budget only. Main funding source(s): FWO - Flanders Research Foundation.

**Background/Introduction:** The international classification of disease (ICD) codes are used worldwide for classification of hospitalisations. The codes are used for administrative, financial and research purposes. It is known however that errors occur in the coding process.

**Purpose:** To investigate methods for automatic classification of disease in unstructured medical records using natural language processing (NLP) and to compare these to conventional ICD-10 coding.

**Methods:** Two datasets were used: the open-source Medical Information Mart for Intensive Care (MIMIC)-III dataset (n=40.000) for algorithm testing and a dataset from a hospital in Belgium (n = 8041). In the Belgian dataset automated pseudonymization was performed prior to further analysis. A training, validation and test set was used in all methods. Automated searches using NLP algorithms were performed for the diagnoses "atrial fibrillation" and "heart failure". The information extraction methods that were used were: rule-based search, logistic regression, term frequency-inverse document frequency (TF-IDF), XGBoost and bio-bidirectional encoder representations from transformers (BioBERT). All algorithms were tested on the MIMIC-III dataset. Precision (or positive predictive value), recall (or sensitivity) and accuracy were calculated for each model. The best performing algorithm was then deployed on the Belgian dataset.

**Results and discussion:** Results for the different NLP models are depicted in Table 1. Non-compatible results in NLP versus ICD-coding often indicated errors in ICD-coding. NLP algorithms could thus be used to improve the ICD-coding process in hospitals.

**Conclusion:** A high accuracy can be achieved when using NLP algorithms for diagnostic classification of patient records. Also, NLP algorithms can be used to identify ICD-coding errors and optimise the ICD-coding process.

	Atrial fibrillation			Heart failure		
	Precision	Recall	Accuracy	Precision	Recall	Accuracy
<b>MIMIC-III</b>						
Rule-based	0.98	0.45	0.72	0.93	0.53	0.75
Logistic regression	0.97	0.89	0.93	0.88	0.84	0.86
TF-IDF	0.94	0.84	0.89	0.86	0.79	0.83
XG-boost on TF-IDF matrix	0.96	0.92	0.94	0.88	0.87	0.87
BioBERT	0.91	0.74	0.84	/	/	/
<b>Belgian dataset</b>						
XG-boost on TF-IDF matrix	0.83	0.85	0.93	0.88	0.95	0.96

Table 1