Article published online: 2023-12-26

# Appendix: Content Summaries of Best Papers for the Natural Language Processing Section of the 2023 IMIA Yearbook

### Ahne A, Khetan V, Tannier X, Rizvi MdIH, Czernichow T, Orchard F, Bour C, Aano A, Fagherazzi G

### Extraction of explicit and implicit cause-effect relationships in patient-reported diabetes-related tweets from 2017 to 2021: Deep learning approach

### JMIR Med Inform 2022;10(7):e37201. doi:10.2196/37201

In this paper, the authors aim at providing a deep learning-based method to extract implicit and explicit relations of cause/effect for diabetes from tweets, and a methodology to understand opinions/feelings reported by patients from a causality perspective. They fine-tuned a BERTweet model on 562,000 tweets annotated with emotion information in order to detect causal sentences. Then, they designed a CRF model using BERTweet features to identify possible cause-effects associations from 265,000 causal sentences. This method allows the authors to obtain several clusters for cause-effect (diabetes, death, insulin), including emotions (anger, fear, sadness) reported by diabetes patients.

### Li Y, Wehbe RM, Ahmad FS, Wang H, Luo Y

### A comparative study of pretrained language models for long clinical text

### J Am Med Inform Assoc 2023 Jan 18;30(2):340-7. doi:10.1093/jamia/ocac225

This paper proposes to enrich Transformers models with clinical knowledge, which allows the authors to achieve state-of-the-art results on biomedical NLP tasks. Nevertheless, the authors highlight that the self-attention mechanism uses a lot of memory and does not allow to process long texts (limitation of 512 sub-units: e.g., discharge summaries from MIMIC have 2,984 tokens on average). They produced two domain-enriched language models based on Longformer (Clinical-Longformer) and BigBird (Clinical-BigBird) to process up to 4,096 sub-units. Those models outperformed existing models (BERT, RoBERTa, BioBERT, and ClinicalBERT) on three tasks (NLI @ medNLI ; QA @ emrQA-relations ; NER @ i2b2 2014). We notice that the source code is available.

### Phatak A, Savage DW, Ohle R, Smith J, Mago V

### Medical Text Simplification Using Reinforcement Learning (TESLEA): Deep Learning-Based Text Simplification Approach

### JMIR Med Inform 2022 Nov 18;10(11):e38095. doi: 10.2196/38095

The authors of this paper highlight that abstracts of scientific papers are publicly available, but they are hard to understand due to the use of medical vocabulary. They develop a text simplification method based on deep-learning trained on 3,568 complex-simple paragraphs (training) and evaluated on 480 paragraphs. Several scores are used to evaluate all aspects: FKGL (Flesch-Kincaid Grade Level), ROUGE, SARI (Simplified Automatic Readability Index), Likert scale. In addition, several examples of generated medical paragraphs are given in the paper, including texts generated by other systems (BART fine-tuned, BART-UL, MUSS, Keep-It-Simple, PEGASUS), which allows to compare all produced outputs.