

# Year 2021: COVID-19, Information Extraction and BERTization among the Hottest Topics in Medical Natural Language Processing

Natalia Grabar<sup>1</sup>, Cyril Grouin<sup>2</sup>, Section Editors for the IMIA Yearbook Section on Natural Language Processing

<sup>1</sup> STL, CNRS, Université de Lille, Domaine du Pont-de-bois, Villeneuve-d'Ascq cedex, France

<sup>2</sup> Université Paris Saclay, CNRS, Laboratoire Interdisciplinaire des Sciences du Numérique, Orsay, France

## Summary

**Objectives:** Analyze the content of publications within the medical natural language processing (NLP) domain in 2021.

**Methods:** Automatic and manual preselection of publications to be reviewed, and selection of the best NLP papers of the year. Analysis of the important issues.

**Results:** Four best papers have been selected in 2021. We also propose an analysis of the content of the NLP publications in 2021, all topics included.

**Conclusion:** The main issues addressed in 2021 are related to the investigation of COVID-related questions and to the further adaptation and use of transformer models. Besides, the trends from the past years continue, such as information extraction and use of information from social networks.

## Keywords

Natural language processing; semi-automatic selection of publication; topics; issues; 2021

Yearb Med Inform 2022;254-61

<http://dx.doi.org/10.1055/s-0042-1742547>

## 1 Introduction

Natural Language Processing (NLP) aims at providing methods, tools and resources designed in order to mine textual and narrative documents, and to make it possible to access the information they convey [1]. While human languages are complex (as an example, learning a human language requires many years in order to be fluent), the importance of using NLP approaches to mine documents produced by humans has been pointed out since a long time [2]. In this synopsis, we first present the selection process applied this year and then we analyze the content of some publications. More particularly, we focus on several important issues such as robustness of the methods, reproducibility of the results, as well as the originality of the research questions addressed in 2021.

## 2 The Selection Process

In order to identify all papers published during the year 2021 in the field of NLP for the biomedical domain, we queried two databases: Meline<sup>1</sup>, specifically dedicated to the biomedical domain, and the Association for Computational Linguistics (ACL) anthology<sup>2</sup>, a database that brings together the major NLP conferences (ACL, COLING, EMNLP, LREC, NAACL, etc.) and journals, since some NLP studies concerning the biomedical domain are published in conferences and journals which are not indexed by PubMed.

<sup>1</sup> <https://pubmed.ncbi.nlm.nih.gov/>

<sup>2</sup> <https://www.aclweb.org/anthology/>

We applied the basic query we defined last year on MEDLINE:

```
(English[LA]
AND journal article[PT]
AND 2021[DP]
AND (medical OR clinical
OR natural)
AND "language processing")
```

all journal papers published in English in 2021, having abstract, and composed of sequences “clinical language processing” or “medical language processing” or “natural language processing”. As of 2022, January 17<sup>th</sup>, we collected 1,204 entries. We applied a similar query on the ACL anthology database and collected 36 entries. In order to process those 1,240 papers, we automatically scored the papers. Indeed, all the candidate papers are not specifically related to the NLP domain despite the use of one of the three sequences from the query. For instance, they may be related to other sections from this Yearbook (Public health, Bioinformatics, Knowledge representation, etc.) and not address major issues of the NLP section. Hence, we applied three sets of rules we defined in 2018 and we reused each year, while identifying best papers in a previous edition, in order to compute global scores for each publication.

The first set of rules is based upon the name of the journal (both full name and concepts found in the name):

- the positive score is assigned to the main journals in which the biomedical NLP research is usually published by the NLP community (Biomedical informatics insights, International journal

of medical informatics, Journal of the American Medical Informatics Association (JAMIA) Journal of Biomedical Informatics (JBI), BMC bioinformatics);

- the negative score is assigned to journals not specifically related to NLP, but to other domains such as Cognitive studies and Communication disorders (Neuroscience, Human brain mapping, Operative neurosurgery, Speech therapy, etc.). We also dismiss survey papers and papers published in the IMIA Yearbook. We manually defined this set of journals in order to rule out those false positives.

The second set of rules relies on concepts found in both the title and abstract of papers:

- the positive score is assigned to concepts typically involved in papers related to NLP. Those concepts may be related to objectives, resources, and tools (such as *natural language processing, NLP, named entity recognition, NER, part of speech, POS, tagged words, semantic, syntax, biomedical entity, meanings, electronic health record, EHR, reports, notes, clinical text, text corpus, free text, unstructured text, tweets, PubMed, Social Media, MedDRA, UMLS, annotated data, Metamap*);
- the negative score is assigned to concepts that are usually involved in studies related to disorders involving anatomical parts or language abilities (such as *word processing, language production, language comprehension, voice quality, posterior superior temporal gyrus, pSTG, posterior superior temporal sulcus, pSTS, inferior fronto-occipital fasciculus, IFOF, dorsolateral prefrontal cortex, cortex, language lateralization, chemical fragment, fragment chemistry, brain structures, verbal intelligence, cerebral, positive mismatch responses, pMMRs, prelingual, postlingual, cochlear, aphasia, SAPS, cortical, language function, infants*).

The third set of rules is also applied on titles and abstracts, and focuses on the concepts describing the methodology used in papers:

- the positive score is assigned to papers using classical NLP methods or evaluation metrics (such as *annotation*

*tool, text-mining, rule-based, regular expression, lexicon, CRF, recall, precision, F1-score, F-measure, accuracy, inter-annotator agreement, Kappa, classify/classifier, detect, extract, extraction, predict, predicting, text simplification, lexical simplification*);

- the negative score is assigned to papers claiming to use the NLP methods, such as pointed out by sequences like *using natural language processing, using NLP, perform a Natural Language Processing analysis*. Such papers are downgraded because the NLP claims are usually limited to the use of the existing and ready-to-use NLP tools while the main contribution of papers is related to the analysis of tool results rather than to the improvements made to the NLP methods and issues. Note that such papers are usually related to other areas from medical informatics: the researchers take advantage of the existing tools.

Contrary to previous years, we also decided to limit the number of journals to be processed in order to dismiss either non-NLP journals (e.g., cognitive issues are not related to natural language processing) and low-quality papers that would have been published in secondary journals. We focused on the top main NLP journals (namely JAMIA, JBI and JBS) since the best work in the NLP domain is generally published in those journals. Nevertheless, since the JAMIA is the journal of a US association, it may have impacted the number of US papers finally kept as candidates.

For each of the 1,240 candidate papers, the final score ranked from 0.05 to 1 (cf. Figure 2). This score has been used as a meta element during the manual selection of the top-13 papers. Indeed, the section

editors did not fully rely on the scores but only used them as additional information. Hence, both section editors independently browsed the abstracts, keywords and automatic scores, and assigned a *Yes / Maybe / No* score to each paper. All papers having at least a *Yes* or *Maybe* score have been kept for the next step of the selection. At this stage, 116 candidate papers remained (i.e., a subset of 9.35% of the whole dataset). We then performed an adjudication process, in order to choose the final 13 candidates to be proofread by external reviewers. We paid attention to the topics addressed by the researchers so as to provide enough diversity. As a result, out of the thirteen papers, nine come from the USA, two from the UK, one from Canada and one from South Korea. This year, we noticed that our process produces an unexpected selection of candidates mainly from English-speaking countries, with an over-representation of papers from the USA. We hypothesize that, in this post-COVID-19 period, researchers from the US maintained their level of submissions while European researchers decreased the number of submitted papers.

Table 1 lists the four best papers of the 2022 IMIA Yearbook for the NLP section.

In the next sections, we present the main issues and approaches addressed in the preselected publications.

### 3 Current Trends in Biomedical NLP

To present the current trends in biomedical NLP observed during this last year, we propose an analysis of 200 top-citations according to the scores computed automatically. First, we analyze the languages addressed

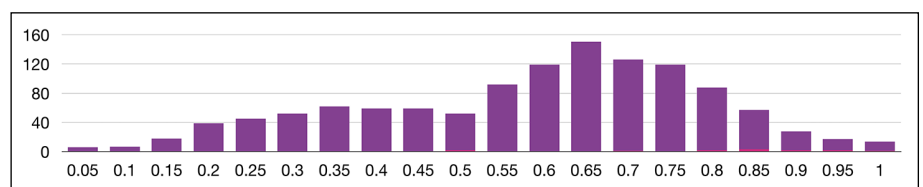


Fig. 2 Distribution of papers according to the filter scores.

**Table 1** Best paper selection for the NLP section of the IMIA Yearbook of Medical Informatics 2022. The articles are listed in alphabetical order of the first author's surname.

Section
<b>Natural Language Processing</b>
<ul style="list-style-type: none"> <li>▪ Flamholz ZN, Crane-Droesch A, Ungar LH, Weissman GE. Word embeddings trained on published case reports are lightweight, effective for clinical tasks, and free of protected health information. <i>J Biomed Inform</i> 2022 Jan;125:103971.</li> <li>▪ Majewska O, Collins C, Baker S, Björne J, Brown SW, Korhonen A, Palmer M. BioVerbNet: a large semantic-syntactic classification of verbs in biomedicine. <i>J Biomed Semantics</i> 2021 Jul 15;12(1):12.</li> <li>▪ Ding X, Mower J, Subramanian D, Cohen T. Augmenting <i>aer2vec</i>: Enriching distributed representations of adverse event report data with orthographic and lexical information. <i>J Biomed Inform</i> 2021 Jul;119:103833.</li> <li>▪ De Angeli K, Gao S, Danciu I, Durbin EB, Wu XC, Stroup A, Doherty J, Schwartz S, Wiggins C, Damesyn M, Coyle L, Penberthy L, Tourassi GD, Yoon HJ. Class imbalance in out-of-distribution datasets: Improving the robustness of the TextCNN for the classification of rare cancer types. <i>J Biomed Inform</i> 2022 Jan;125:103957.</li> </ul>

(Section 3.1) in this year publications. Second, we focus on the topics studied by researchers. For this, we first analyze the keywords provided by the authors (Section 3.2). Then, we group our observations along two lines: main approaches and methods used (Sections 3.3 and 3.4), and some frequent topics studied (Sections 3.5 and 3.6).

### 3.1 Languages Addressed

We can observe that, in 2021, researchers mainly concentrated on studying English-language data. Indeed, compared to previous years, a fewer number of languages were covered: Chinese [3-10], Dutch [11], French [12, 13], Italian [14-16], Japanese [17], Korean [18, 19], Norwegian [20], and Spanish [21-13]. Besides, except for Chinese, there were also very few works done for the languages represented in publications.

We assume that the reasons leading to this situation may be due to the topics addressed and approaches used: (1) for some topics, like COVID-19 or clinical trials, the data are mainly available in English and are very limited in other languages, and (2) for some approaches, like those related to language models, the necessary volume of the data can be available in English only. The issue related to the availability of large and annotated language data in different languages is yet the stumbling block for NLP.

### 3.2 Keywords

We studied the distribution of keywords used to index those 200 top papers. This distribution gives us a global overview of what is important for authors since they chose such keywords to index their papers. We give here the number of papers for which the keyword has been used in order to highlight the amount of papers for which this keyword is relevant. To rephrase it, other papers certainly used those techniques or domains but their authors did not consider it important to use those keywords. We thus observed the following trends for 2021:

- more than half of those papers (123 papers) propose the generic “natural language processing” keyword or its more specialized one “clinical natural language processing”, meaning that their work still relies on NLP-based approaches, rather than “artificial intelligence” (10 papers) which is not used by the original NLP community but has been adopted by *data scientists*;
- “machine-learning” techniques are still more used (30 papers) than “deep learning” techniques (21 papers) or approaches specifically based on “CNN” and other “neural networks” (4 papers); “word embeddings” and “contextualized embeddings” are also central for a significant number of papers (12 papers); the use of “BERT” and other “transformer” models is also important for some other studies (9 papers); we also noticed a few papers

focusing on strategies to improve results, such as “data augmentation” (4 papers), “ensemble” techniques (3 papers), and “feature selection” (2 papers);

- several authors also present keywords to highlight the content they used to perform their work: this year, we noticed slightly more papers based on “electronic health records” or “clinical notes” or “clinical texts” or “radiology reports” (31 papers) than “social media” or “twitter” (19 papers); beyond the more complex access to EHRs, the need for pharmacovigilance in these pandemic times makes it mandatory to search for patient testimony;
- out of all tasks addressed in those papers, we noticed the following ones that are still widely processed: “named entity recognition” (11 papers), “information extraction” (8 papers), “topic modeling” (3 papers), “text classification” (3 papers), “relation extraction” (3 papers), “sentiment analysis” (2 papers), “event extraction” (2 papers), and “clustering” (2 papers);
- in terms of thematic, the first subject still concerns “COVID-19” or “coronavirus” studies (18 papers), including for “pharmacovigilance”, “epidemiology”, “infodemiology” or “infoveillance” purposes (11 papers); other studies focus on specific diseases, such as “cancer”, “cardiovascular disease”, “dementia”, “heart failure”, “mental health”, “pain”, “pulmonary embolism”, “stroke” (2 papers for each keyword); more specifically, the identification and/or processing of “social determinants of health” (3 papers) as well as “adverse drug event” and “adverse drug reaction” (4 papers) were also considered; we also observed an insight on specific clinical services in those studies: “emergency department”, “ophthalmology”, and “radiology” (all of them for less than 4 papers). At last, a specific need for “evidence-based medicine” (3 papers) also occurred.

Even if those keywords give a global view of work done in these papers, they only reflect a small and general part of the work that has been done by the authors. Indeed, keywords are essentially used for indexing papers—whether those keywords are free or required from a given terminology—while

we used them to draw an overview of the research done last year. Nevertheless, the trends observed within the set of keywords are also reflected in the analysis provided in the following sections.

### 3.3 Information Extraction

The purpose of information extraction is to localize directly within textual unstructured documents specific pieces of information, like mentions of patients, their disorders, procedures, drugs, adverse events, various relations, etc. Thanks to many years of research and several NLP challenges [24-28], information extraction is now one of the NLP areas which currently delivers reliable and exploitable results, which of course requires that the methods are first adapted and fitted to a given research problem. Hence, the information extraction methods are quite widely used in various clinical contexts for extracting different kinds of entities and relations, such as:

- detection of events [29, 30], including self-harm events [31];
- extraction of diagnoses [13, 32-35] and their codes [36-38];
- recognition of named entities [5, 14, 39-41], and more specifically of personal information [21, 42, 43] and family history [20];
- localization of advices [44] and arguments [45] in scientific literature;
- extraction of relations [46-48], including temporal [49] and causality [50, 51] relations.

Recruitment of patients for clinical trials is one of the research questions which heavily relies on information extraction methods. Indeed, the inclusion and exclusion criteria require to consider various aspects related to patients, their history, habits, treatment, procedures, etc. In 2021, several publications were dedicated to this research question [52-54], thus following the tendency from past years.

### 3.4 Language Models

Language models group together a set of methods based on word embeddings. More specifically, transformer-based models are being researched more intensively: BERT

[55], BioBERT [56], FlauBERT [57], etc. These models indeed provide excellent results for different NLP tasks and applications. This may explain their heavy exploitation in a majority of publications cited in this chapter. We indeed witness a kind of BERTization of the NLP area: to obtain competitive results, the exploitation of the BERT-issued models is indicated. Yet, such models need high computation resources and large annotated corpora to train new models. To overcome these limitations, several alternative solutions have been addressed this year by researchers: fine-tuning of existing models [58-63], domain adaptation [43, 64], transfer learning [48, 50, 60], self-training [43]. To go further in these directions, new trends were also observed in 2021, like reuse of older architectures based on fastText [65] and word2vec [66] enriched with basic language information: orthographic and lexical [67], syntactic-semantic classes [68], medical knowledge [46], subword embeddings [14, 69], vector retrofitting [67, 68].

Note also that multi-task systems were also proposed and can satisfy several NLP tasks adapted to the medical area [41].

### 3.5 COVID-19

Due to the recent pandemic situation, COVID-19 remains an important issue addressed by researchers in 2021. In previous years, the main source of data was represented by social networks: this kind of data was emerging, and the researchers had to exploit population-created contents. This year, we can observe a new tendency, which consists of exploiting different sources of data: social networks, scientific literature, and also clinical data. Hence, this year, the sources available cover all the aspects of the COVID-19 cycle: clinical observations, scientific research, and population experience. This indicates that the data on COVID-19 are becoming more widely available and, more specifically, are also describing a larger number of patients with COVID-19, especially in hospitals. This permits a more systematic and comprehensive study of aspects related to COVID-19, such as:

- extraction of diagnoses and symptoms related to COVID-19 from clinical text [13, 34, 35, 64];

- link between COVID-19 and mental health [70];
- creation of a symptom lexicon on sequelae due to COVID-19 from clinical notes [71];
- preparing clinical trials for COVID-19 vaccination [53];
- generation of dialogues on COVID-19 to help population in a better understanding of inherent issues [4],
- finding advices in the scientific literature on different aspects related to COVID-19 in order to help the population get timely answers to their questions [44];
- detection of disinformation about COVID-19 vaccines in social media [72],
- a scoping review on the analysis of the use of AI methods applied in the COVID-19 research [73].

### 3.6 Mental Health

The processing and study of neurological and psychiatric disorders remains another important topic, which has been investigated by researchers for several years. This year, we can notice several issues:

- improvement of the diagnosis of this kind of disorders thanks to NLP methods [74];
- identification of the temporal evolution and stage of mental disorders in order to improve their treatment [16];
- detection of self-harm and suicidal data [31, 75];
- link of mental health with other disorders, like COVID-19 [70, 76], HIV [77], cancers [78], or drug abuse [79];
- analysis of psychedelic session narratives in order to predict changes in substance use [80];
- creation of online support systems for people suffering from mental disorders [78, 81].

## 4 Conclusion

NLP researchers are evolving in accordance with the time they are living in, which is mainly reflected by two aspects in 2021: (1) from the methodological point of view, the researchers are using methods that proved their efficiency for different NLP tasks, including within the clinical context, such in-

formation extraction methods for localizing and extracting precise pieces of information in medical documents or transformer-based approaches for improving the overall results thanks to a better representation of textual data; (2) from the thematic point of view, the issues related to COVID-19, mental health, oncology, etc. are attractive and timely for the research. While in previous years there was a great deal of variability in the languages addressed by the researchers, this variability is reduced in 2021. We assume this is due to the little textual data available for some topics and languages. In this context, the researchers worked on available annotated datasets in English. Some additional time will be necessary to create and enrich similar datasets in other languages.

For future researches, and especially for papers describing experiments using BERT-based models, we hope that the authors will focus more on analyzing the results than on providing basic discussions about computed results that range between 0.95 and 1.00, regardless of the metric used. Indeed, if current publications mainly rely on quantitative results, they lack constructive perspectives and qualitative insights. Deep learning must proceed together with an in-depth analysis of the obtained outcomes! In addition, knowledge must not be discarded in NLP-based research and algorithms must not be the only solution.

## References

- Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. *J Am Med Inform Assoc* 2011 Sep-Oct;18(5):544-51.
- Friedman C, Hripcsak G. Natural language processing and its future in medicine. *Acad Med* 1999 Aug;74(8):890-5.
- Li J, Zhong S, Chen K. MLEC-QA: A Chinese Multi-Choice Biomedical Question Answering Dataset. In: *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. ACL*; 2021. p. 8862-74.
- Zhou M, Li Z, Tan B, Zeng G, Yang W, He X, et al. On the generation of medical dialogs for COVID-19. In: *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing 2021, Volume 2: Short Papers. ACL*; 2021. p. 886-96.
- Kong J, Zhang L, Jiang M, Liu T. Incorporating multi-level CNN and attention mechanism for Chinese clinical named entity recognition. *J Biomed Inform* 2021 Apr;116:103737.
- Jia Q, Zhang D, Yang S, Xia C, Shi Y, Tao H, et al. Traditional Chinese medicine symptom normalization approach leveraging hierarchical semantic information and text matching with attention mechanism. *J Biomed Inform* 2021 Apr;116:103718.
- Wu Z, Liang J, Zhang Z, Lei J. Exploration of text matching methods in Chinese disease Q&A systems: A method using ensemble based on BERT and boosted tree models. *J Biomed Inform* 2021 Mar;115:103683.
- Zhou L, Liu S, Li C, Sun Y, Zhang Y, Li Y, et al. Natural Language Processing Algorithms for Normalizing Expressions of Synonymous Symptoms in Traditional Chinese Medicine. *Evid Based Complement Alternat Med* 2021 Oct 11;2021:6676607.
- Xu Z, Xu Y, Cheung F, Cheng M, Lung D, Law YW, et al. Detecting suicide risk using knowledge-aware natural language processing and counseling service data. *Soc Sci Med* 2021 Aug;283:114176.
- Shen S, Zhu C, Fan C, Wu C, Huang X, Zhou L. Research on the evolution and driving forces of the manufacturing industry during the "13th five-year plan" period in Jiangsu province of China based on natural language processing. *PLoS One* 2021 Aug 18;16(8):e0256162.
- Nobel JM, Puts S, Weiss J, Aerts HJWL, Mak RH, Robben SGF, et al. T-staging pulmonary oncology from radiological reports using natural language processing: translating into a multi-language setting. *Insights Imaging* 2021 Jun 10;12(1):77.
- Wajsbürt P, Sarfati A, Tannier X. Medical concept normalization in French using multilingual terminologies and contextual embeddings. *J Biomed Inform* 2021 Feb;114:103684.
- Ferté T, Cossin S, Schaeverbeke T, Barnetche T, Jouhet V, Hejblum BP. Automatic phenotyping of electronic health record: PheVis algorithm. *J Biomed Inform* 2021 May;117:103746.
- Lauriola I, Aiolfi F, Lavelli A, Rinaldi F. Learning adaptive representations for entity recognition in the biomedical domain. *J Biomed Semantics* 2021 May 17;12(1):10.
- Hammami L, Paglialonga A, Pruneri G, Torresani M, Sant M, Bono C, et al. Automated classification of cancer morphology from Italian pathology reports using Natural Language Processing techniques: A rule-based approach. *J Biomed Inform* 2021 Apr;116:103712.
- Viani N, Botelle R, Kerwin J, Yin L, Patel R, Stewart R, et al. A natural language processing approach for identifying temporal disease onset information from mental healthcare text. *Sci Rep* 2021 Jan 12;11(1):757.
- Matsuda S, Ohtomo T, Tomizawa S, Miyano Y, Mogi M, Kuriki H, et al. Incorporating Unstructured Patient Narratives and Health Insurance Claims Data in Pharmacovigilance: Natural Language Processing Analysis of Patient-Generated Texts About Systemic Lupus Erythematosus. *JMIR Public Health Surveill* 2021 Jun 29;7(6):e29238.
- Shin D, Kam HJ, Jeon MS, Kim HY. Automatic Classification of Thyroid Findings Using Static and Contextualized Ensemble Natural Language Processing Systems: Development Study. *JMIR Med Inform* 2021 Sep 21;9(9):e30223.
- Kim D, Oh J, Im H, Yoon M, Park J, Lee J. Automatic Classification of the Korean Triage Acuity Scale in Simulated Emergency Rooms Using Speech Recognition and Natural Language Processing: a Proof of Concept Study. *J Korean Med Sci* 2021 Jul 12;36(27):e175.
- Brekke PH, Rama T, Pilán I, Nytrø Ø, Øvrelid L. Synthetic data for annotation and extraction of family history information from clinical text. *J Biomed Semantics* 2021 Jul 14;12(1):11.
- Pérez-Díez I, Pérez-Moraga R, López-Cerdán A, Salinas-Serrano JM, la Iglesia-Vayá M. De-identifying Spanish medical texts - named entity recognition applied to radiology reports. *J Biomed Semantics* 2021 Mar 29;12(1):6.
- Campillos-Llanos L, Valverde-Mateos A, Capllonch-Carrión A, Moreno-Sandoval A. A clinical trials corpus annotated with UMLS entities to enhance the access to evidence-based medicine. *BMC Med Inform Decis Mak* 2021 Feb 22;21(1):69. Erratum in: *BMC Med Inform Decis Mak* 2021 Apr 7;21(1):118.
- Villena F, Pérez J, Lagos R, Dunstan J. Supporting the classification of patients in public hospitals in Chile by designing, deploying and validating a system based on natural language processing. *BMC Med Inform Decis Mak* 2021 Jul 1;21(1):208. Erratum in: *BMC Med Inform Decis Mak* 2021 Jul 20;21(1):220.
- Uzuner Ö. Second i2b2 workshop on natural language processing challenges for clinical records. *AMIA Annu Symp Proc* 2008 Nov 6:1252-3.
- Uzuner Ö, South BR, Shen S, DuVall SL. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. *J Am Med Inform Assoc* 2011 Sep-Oct;18(5):552-6.
- UzZaman N, Llorens H, Derczynski L, Allen J, Verhagen M, Pustejovsky J. Semeval-2013 task 1: Tempeval-3: Evaluating time expressions, events, and temporal relations. In *Second Joint Conference on Lexical and Computational Semantics (\*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*; 2013. p. 1-9.
- Sun W, Rumshisky A, Uzuner O. Evaluating temporal relations in clinical text: 2012 i2b2 Challenge. *J Am Med Inform Assoc* 2013 Sep-Oct;20(5):806-13.
- Henry S, Wang Y, Shen F, Uzuner O. The 2019 National Natural language processing (NLP) Clinical Challenges (n2c2)/Open Health NLP (OHNLP) shared task on clinical concept normalization for clinical records. *J Am Med Inform Assoc* 2020 Oct 1;27(10):1529-37. Erratum in: *J Am Med Inform Assoc* 2021 Oct 12;28(11):2546.
- Naik A, Lehman JF, Rose C. Adapting event extractors to medical data: Bridging the covariate shift. In: *Proc of the 16th Conf of the European Chapter of the Association for Computational Linguistics: Main Volume. ACL*; 2021. p. 2963-75.
- Magge A, Tutubalina E, Miftahutdinov Z, Alimova I, Dirksen A, Verberne S, et al. DeepADEMiner: a deep learning pharmacovigilance pipeline for extraction and normalization of adverse drug event mentions on Twitter. *J Am Med Inform Assoc* 2021 Sep 18;28(10):2184-92.
- Rozova V, Witt K, Robinson J, Li Y, Verspoor K. Detection of self-harm and suicidal ideation in

- emergency department triage notes. *J Am Med Inform Assoc* 2022 Jan 29;29(3):472-80.
32. Haoran W, Chen W, Xu S, Xu B. Counterfactual supporting facts extraction for explainable medical record based diagnosis with graph network. In: Proc of the 2021 Conf of the North American Chapter of the Assoc for Computational Linguistics: Human Language Technologies. ACL; 2021. p. 1942–55.
  33. Amiri H, Mohtarami M, Kohane I. Attentive multi-view text representation for differential diagnosis. In: Proc of the 59th Ann Meeting of the Ass for Comp Linguistics and the 11th Inter Joint Conf on Natural Language Processing. ACL; 2021. p. 1012–9.
  34. Lybarger K, Ostendorf M, Thompson M, Yetsigen M. Extracting COVID-19 diagnoses and symptoms from clinical text: A new annotated corpus and neural event extraction framework. *J Biomed Inform* 2021 May;117:103761.
  35. Zhao J, Grabowska ME, Kerchberger VE, Smith JC, Eken HN, Feng Q, et al. ConceptWAS: A high-throughput method for early identification of COVID-19 presenting symptoms and characteristics from clinical notes. *J Biomed Inform* 2021 May;117:103748.
  36. Zhou T, Cao P, Chen Y, Liu K, Zhao J, Niu K, et al. Automatic ICD coding via interactive shared representation networks with self-distillation mechanism. In: Proc of the 59th Ann Meeting of the Assoc for Comp Linguistics and the 11th Inter Joint Conf on Natural Language Processing. ACL; 2021. p. 5948–57.
  37. Liu Y, Cheng H, Klopfer R, Gormley MR, Schaaf T. Effective convolutional attention network for multi-label clinical document classification. In: Proc of the 2021 Conf on Empirical Methods in Natural Language Processing. ACL; 2021. p. 5941–53.
  38. Dong H, Suárez-Paniagua V, Whiteley W, Wu H. Explainable automated coding of clinical notes using hierarchical label-wise attention networks and label embedding initialisation. *J Biomed Inform* 2021 Apr;116:103728.
  39. Zhou B, Cai X, Zhang Y, Yuan X. An end-to-end progressive multi-task learning framework for medical named entity recognition and normalization. In: Proc of the 59th Ann Meeting of the Assoc for Comp Linguistics and the 11th Inter Joint Conf on Natural Language Processing. ACL; 2021. p. 6214–24.
  40. Vashishth S, Newman-Griffis D, Joshi R, Dutt R, Rosé CP. Improving broad-coverage medical entity linking with semantic type prediction and large-scale datasets. *J Biomed Inform* 2021 Sep;121:103880.
  41. Mulyar A, Uzuner Ö, McInnes B. MT-clinical BERT: scaling clinical information extraction with multitask learning. *J Am Med Inform Assoc* 2021 Sep 18;28(10):2108-15.
  42. Valizadeh M, Ranjbar-Noiey P, Caragea C, Parde N. Identifying medical self-disclosure in online communities. In: Proc of the 2021 Conf of the North American Chapter of the Association for Comp Linguistics: Human Language Technologies. ACL; 2021. p. 4398–408.
  43. Liao S, Kiros J, Chen J, Zhang Z, Chen T. Improving domain adaptation in de-identification of electronic health records through self-training. *J Am Med Inform Assoc* 2021 Sep 18;28(10):2093-100.
  44. Li Y, Wang J, Yu B. Detecting health advice in medical research literature. In: Proc of the 2021 Conf on Empirical Methods in Natural Language Processing. ACL; 2021. p. 6018–29.
  45. Stylianou N, Vlahavas I. TransforMED: End-to-end Transformers for Evidence-Based Medicine and Argument Mining in medical literature. *J Biomed Inform* 2021 May;117:103767.
  46. Roy A, Pan S. Incorporating medical knowledge in BERT for clinical relation extraction. In: Proc of the 2021 Conf on Empirical Methods in Natural Language Processing. ACL; 2021. p. 5357–66.
  47. Kanjirang V, Rinaldi F. Enhancing Biomedical Relation Extraction with Transformer Models using Shortest Dependency Path Features and Triplet Information. *J Biomed Inform* 2021 Oct;122:103893.
  48. Legrand J, Toussaint Y, Raïssi C, Coulet A. Syntax-based transfer learning for the task of biomedical relation extraction. *J Biomed Semantics* 2021 Aug 18;12(1):16.
  49. Alfattni G, Peek N, Nenadic G. Attention-based bidirectional long short-term memory networks for extracting temporal relationships from clinical discharge summaries. *J Biomed Inform* 2021 Nov;123:103915.
  50. Hussain M, Satti FA, Hussain J, Ali T, Ali SI, Bilal HSM, et al. A practical approach towards causality mining in clinical text using active transfer learning. *J Biomed Inform* 2021 Nov;123:103932.
  51. Ma X, Imai T, Shinohara E, Kasai S, Kato K, Kagawa R, et al. EHR2CCAS: A framework for mapping EHR to disease knowledge presenting causal chain of disorders - chronic kidney disease example. *J Biomed Inform* 2021 Mar;115:103692.
  52. Percha B, Pisapati K, Gao C, Schmidt H. Natural language inference for curation of structured clinical registries from unstructured text. *J Am Med Inform Assoc* 2021 Dec 28;29(1):97-108.
  53. Du J, Wang Q, Wang J, Ramesh P, Xiang Y, Jiang X, et al. COVID-19 trial graph: a linked graph for COVID-19 clinical trials. *J Am Med Inform Assoc* 2021 Aug 13;28(9):1964-9.
  54. Liu H, Chi Y, Butler A, Sun Y, Weng C. A knowledge base of clinical trial eligibility criteria. *J Biomed Inform* 2021 May;117:103771.
  55. Devlin J, Chang M-W, Lee K, Kristina Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. In: Association for Computational Linguistics, editor. Proc of NAACL-HLT 2019. Minneapolis, Minnesota; 2019. p. 4171–86.
  56. Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* 2020 Feb 15;36(4):1234-40.
  57. Le H, Vial L, Frej J, Segonne V, Coavoux M, Lecouteux B, et al. FlauBERT: Unsupervised language model pre-training for French. In: Proc of the 12th Language Resources and Evaluation Conf. Marseille, France: European Language Resources Association; 2020. p. 2479-90.
  58. Flamholz ZN, Crane-Droesch A, Ungar LH, Weissman GE. Word embeddings trained on published case reports are lightweight, effective for clinical tasks, and free of protected health information. *J Biomed Inform* 2022 Jan;125:103971.
  59. Noh J, Kavuluru R. Improved biomedical word embeddings in the transformer era. *J Biomed Inform* 2021 Aug;120:103867.
  60. Bear Don't Walk Iv OJ, Sun T, Perotte A, Elhadad N. Clinically relevant pretraining is all you need. *J Am Med Inform Assoc* 2021 Aug 13;28(9):1970-6.
  61. Michalopoulos G, Wang Y, Kaka H, Chen H, Wong A. UmlsBERT: Clinical domain knowledge augmentation of contextual embeddings using the Unified Medical Language System Metathesaurus. In: Proc of the 2021 Conf of the North American Chapter of the Ass for Comp Linguistics: Human Language Technologies. ACL; 2021. p. 1744–53.
  62. Liu F, Shareghi E, Meng Z, Basaldella M, Collier N. Self-alignment pretraining for biomedical entity representations. In: Association for Computational Linguistics, editor. Proc of the 2021 Conf of the North American Chapter of the Ass for Comp Linguistics: Human Language Technologies. 2021. p. 4228–38.
  63. Amir S, van de Meent J-W, Wallace BC. On the impact of random seeds on the fairness of clinical classifiers. In: Association for Computational Linguistics, editor. Proc of the 2021 Conf of the North American Chapter of the Ass for Comp Linguistics: Human Language Technologies. 2021. p. 3808–23.
  64. Wang J, Abu-El-Rub N, Gray J, Pham HA, Zhou Y, Manion FJ, et al. COVID-19 SignSym: a fast adaptation of a general clinical NLP tool to identify and normalize COVID-19 signs and symptoms to OMOP common data model. *J Am Med Inform Assoc* 2021 Jun 12;28(6):1275-83.
  65. Bojanowski P, Grave E, Joulin A, Tomas Mikolov T. Enriching word vectors with subword information. *Trans Assoc Comput Linguist* 2017; 5(1):135–46.
  66. Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionality. In: Proc of the 26th International Conf on Neural Information Processing Systems – Volume 2. 2012. p. 3111-9.
  67. Ding X, Mower J, Subramanian D, Cohen T. Augmenting aer2vec: Enriching distributed representations of adverse event report data with orthographic and lexical information. *J Biomed Inform* 2021 Jul;119:103833.
  68. Majewska O, Collins C, Baker S, Björne J, Brown SW, Korhonen A, et al. BioVerbNet: a large semantic-syntactic classification of verbs in biomedicine. *J Biomed Semantics* 2021 Jul 15;12(1):12.
  69. Kim T, Han SW, Kang M, Lee SH, Kim JH, Joo HJ, et al. Similarity-Based Unsupervised Spelling Correction Using BioWordVec: Development and Usability Study of Bacterial Culture and Antimicrobial Susceptibility Reports. *JMIR Med Inform* 2021 Feb 22;9(2):e25530.
  70. Pachamanova D, Glover W, Li Z, Docktor M, Gujral N. Identifying patterns in administrative tasks through structural topic modeling: A study of task definitions, prevalence, and shifts in a mental health practice's operations during the COVID-19 pandemic. *J Am Med Inform Assoc* 2021 Nov 25;28(12):2707-15.
  71. Wang L, Foer D, MacPhaul E, Lo YC, Bates DW, Zhou L. PASCLeX: A comprehensive post-acute sequelae of COVID-19 (PASC) symptom lexicon

- derived from electronic health record clinical notes. *J Biomed Inform* 2022 Jan;125:103951.
72. Weinzierl MA, Harabagiu SM. Automatic detection of COVID-19 vaccine misinformation with graph link prediction. *J Biomed Inform* 2021 Dec;124:103955.
  73. Guo Y, Zhang Y, Lyu T, Prospero M, Wang F, Xu H, et al. The application of artificial intelligence and data integration in COVID-19 studies: a scoping review. *J Am Med Inform Assoc* 2021 Aug 13;28(9):2050-67.
  74. Shiner B, Levis M, Dufort VM, Patterson OV, Watts BV, DuVall SL, et al. Improvements to PTSD quality metrics with natural language processing. *J Eval Clin Pract* 2021 May 24.
  75. Cliffe C, Seyedsalehi A, Vardavoulia K, Bittar A, Velupillai S, Shetty H, et al. Using natural language processing to extract self-harm and suicidality data from a clinical sample of patients with eating disorders: a retrospective cohort study. *BMJ Open* 2021 Dec 31;11(12):e053808.
  76. Patel R, Smeraldi F, Abdollahyan M, Irving J, Bessant C. Analysis of mental and physical disorders associated with COVID-19 in online health forums: a natural language processing study. *BMJ Open* 2021 Nov 5;11(11):e056601.
  77. Ridgway JP, Uvin A, Schmitt J, Oliwa T, Almirol E, Devlin S, et al. Natural Language Processing of Clinical Notes to Identify Mental Illness and Substance Use Among People Living with HIV: Retrospective Cohort Study. *JMIR Med Inform* 2021 Mar 10;9(3):e23456.
  78. Leung YW, Wouterloot E, Adikari A, Hirst G, de Silva D, Wong J, Bender JL, et al. Natural Language Processing-Based Virtual Cofacilitator for Online Cancer Support Groups: Protocol for an Algorithm Development and Validation Study. *JMIR Res Protoc* 2021 Jan 7;10(1):e21453.
  79. Wright AP, Jones CM, Chau DH, Matthew Glad-den R, Sumner SA. Detection of emerging drugs involved in overdose via diachronic word embeddings of substances discussed on social media. *J Biomed Inform* 2021 Jul;119:103824.
  80. Cox DJ, Garcia-Romeu A, Johnson MW. Predicting changes in substance use following psychedelic experiences: natural language processing of psychedelic session narratives. *Am J Drug Alcohol Abuse* 2021 Jul 4;47(4):444-54.
  81. Hassan A, Ali MDI, Ahammed R, Bourouis S, Khan MM. Development of NLP-Integrated Intelligent Web System for E-Mental Health. *Comput Math Methods Med* 2021 Dec 13;2021:1546343.

**Correspondence to:**

Natalia Grabar  
 STL, CNRS, Université de Lille  
 Domaine du Pont-de-bois  
 59653 Villeneuve-d'Ascq cedex, France  
 E-mail: natalia.grabar@univ-lille.fr