Towards Providing Clinical Insights on Long Covid from Twitter Data

Rohan Bhambhoria⋆[1], Jad Saab⋆[2], Sara Uppal[2], Xin Li[3,7], Artur Yakimovich[4], Junaid Bhatti[5], Nirma Khatri Valdamudi[6], Diana Moyano[7], Michael Bales[8], Elham Dolatabadi[7], and Sedef Akinli Kocak[7]

[1] Queen's University. Kingston, ON, Canada
r.bhambhoria@queensu.ca
[2] Telus Communications Inc. Vancouver, BC, Canada
{jad.saab, sara.uppal}@telus.com
[3] University of Toronto. Toronto, ON, Canada
nix.li@mail.utoronto.ca
[4] Roche Products Ltd. Welwyn Garden City, UK
artur.yakimovich@roche.com
[5] Manulife. Toronto, ON, Canada
junaid_bhatti@manulife.com
[6] University of British Columbia. Vancouver, BC, Canada
nirma.vadlamudi@ubc.ca
[7] Vector Institute. Toronto, ON, Canada
{diana.moyano, elham.dolatabadi, sedef.kocak}@vectorinstitute.ai
[8] Hoffmann-La Roche Ltd. Mississauga, ON, Canada
michael.bales@roche.com

Abstract. From the outset of the COVID-19 pandemic, social media has provided a platform for sharing and discussing experiences in real time. This rich source of information may also prove useful to researchers for uncovering evolving insights into post-acute sequelae of SARS-CoV-2 (PACS), commonly referred to as Long COVID. In order to leverage social media data, we propose using entity-extraction methods for providing clinical insights prior to defining subsequent downstream tasks. In this work, we address the gap between state-of-the-art entity recognition models and the extraction of clinically relevant entities which may be useful to provide explanations for gaining relevant insights from Twitter data. We then propose an approach to bridge the gap by utilizing existing configurable tools, and datasets to enhance the capabilities of these models. Code for this work is available here⋆⋆.

Keywords: Interpretability · Entity Extraction · Healthcare

1 Introduction

Since the emergence of the SARS-CoV-2 virus in late 2019, the resulting COVID-19 pandemic has brought many challenges for patients, healthcare professionals,

⋆ Equal contributions, listed alphabetically
⋆⋆ https://github.com/VectorInstitute/ProjectLongCovid-NER

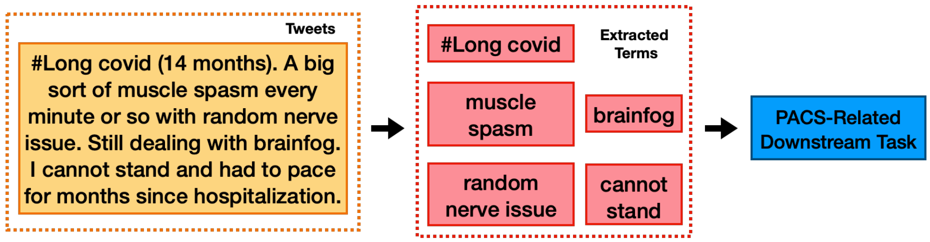

Fig. 1. Our proposed framework for providing insights through entity extraction for Long COVID-Related tasks.

and society. At onset, one particular challenge was a lack of knowledge about both acute and long-term symptoms from infection, and appropriate means of treating or managing them. Before research publications emerged, clinicians had to rely on anecdotal information to guide patient care decisions [27]. Post-acute sequelae of SARS-CoV-2, or Long COVID, describes a condition in which patients have symptoms persisting or recurring for weeks or months following acute COVID-19 infection [7]. Currently, the risk factors leading to the persistence, late onset, or recurrence of symptoms in previously infected COVID-19 patients are unclear [23]. Given the SARS-COV-2 virus is now widely believed to be endemic to the global population [15], it is important to understand the broader impact of the disease, beyond acute infection, and appropriate therapeutic approaches to optimally manage the various patient sub-populations and their disease profiles.

Social media is often turned to by patients as an outlet to express their experience with illness [21]. As such, the authors propose social media may be a useful aid for researchers and clinicians in gaining novel insights into clinical characteristics arising due to emergent illnesses. In this work, we explore the utility of publicly-available, self-reported, user-generated conversations on social media towards capturing terms related to Long COVID symptoms, recoveries and experiences. These terms can then guide investigation into clinical databases, thereby accelerating the discovery, study, description, and hopefully treatment of unexpected consequences from COVID-19. With a few adjustments, the same data processing pipeline may be extensible to other applications, such as emerging infectious diseases or rare and neglected diseases.

Adapting and extending natural language processing (NLP) techniques to extract patient experiences, including symptom evolution and treatment approaches, from user generated data can open doors to better research and characterize the Long COVID phenomenon (Figure 1). However, health language processing with social media is challenging due to the nature of user generated data, such as informal language, short context, and content that is noisy and sparse [13]. To respond, this study aims to utilize deep-learning based entity extraction in order to facilitate decision-making of crucial importance to clinicians, whilst also being interpretable. In this study we also incorporate Metathesaurus, a repository of inter-related biomedical concepts from the Unified Medical Lan-

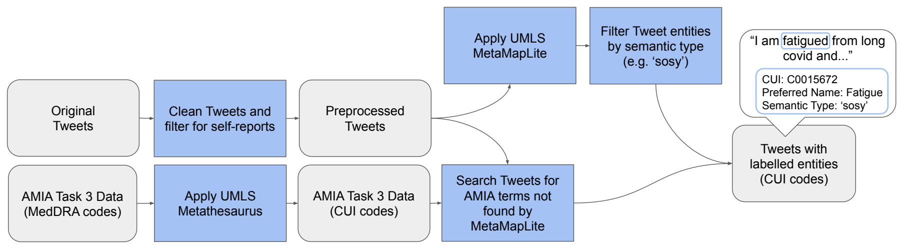

Fig. 2. Pipeline used to label entities with CUI codes.

guage System (UMLS) [3], and our own rule-based matching systems from publicly available datasets in our entity extraction pipeline to enhance its capability.

In this work, our contributions are three-fold:

- First, we propose utilizing entity-extraction methods to provide insights into Long COVID experiences as expressed by patients.
- Secondly, we empirically show that the performance capabilities of state-of-the-art models on existing datasets fall short of human evaluation for relevant Long COVID-related terms. We then propose using MetaMapLite and existing datasets as a way of bridging this gap.
- Finally, we release a dataset and discuss methods which may be utilized in future experiments for augmentation to enable improved extraction of entities from social media text using configurable tools and public datasets. We also release a human annotated test set to serve as a benchmark for this task.

2 Related Work

2.1 Social Media Platform for COVID-19

Analysis of social media data rapidly became popular among researchers and epidemiologists to identify and detect outbreaks of infectious diseases and to interpret public attitudes, behaviours, and perceptions. A systematic review by Tsao and colleagues [25] identified various categories of Twitter use for health research such as surveillance and monitoring [9, 17], and disease control [28]. Among the selected studies in the review, Twitter was the leading social media platform to explore multiple facets of public health research. Mackey and colleagues [11] employed machine learning approaches to investigate COVID-19-related symptoms, experiences with access to testing, and mentions of disease recovery from 4.5 million tweets that were related to COVID-19. Another study conducted by Guo and colleagues [8] was able to extract all the symptoms suggested by Centers for Disease Control and Prevention (CDC) for COVID-19 screening in March, April, and May from 30,732 unique tweets. The findings of this study revealed that mining social media is a promising approach to identify COVID-19-related symptoms earlier than the announcement by the CDC. Motivated by the crucial role of social media in public health surveillance, this study explores mining self-reported experiences from Twitter to identify symptoms related to the "long-haulers," in response to the existing scientific and clinical knowledge gap and many unknowns surrounding Long COVID [1].

2.2 Clinical Information Extraction

To date, various NLP models have been developed to automate extraction of clinical information from both healthcare data and user generated content. In this regard, several challenges and shared tasks have been organized to establish state-of-the-art benchmarks and advance explorations on important information extraction problems in NLP such as entity extraction and normalization from clinical notes and social media posts. In recent years, top performing models in these competitions used transformer architectures for downstream NLP tasks [26, 24, 16, 2]. For instance, in the SMM4H 2021 [12] shared tasks which were organized to promote the use of social media such as Twitter data (tweets) for health applications, the percentage of teams that used either BERT or its variants such as RoBERTA for most of the subtasks was nearly 100%. The dramatic success of transformer models for rich contextual representations led to the creation of various domain specific BERT models such as COVID-Twitter-BERT (CT-BERT) [14] which was trained on a large corpus of tweets about the coronavirus. The model was shown to outperform original BERT-LARGE model on five different Twitter-based datasets. Another variant of the BERT model called UmlsBERT [10] incorporates UMLS Metathesaurus in the pre-training phase in order to build 'semantically enriched' contextual representations that benefit from both the contextual learning (BERT architecture) and the domain knowledge (UMLS Metathesaurus). Given the promising performance of UmlsBERT on clinical NLP tasks, we extend its use to extract clinical terms from social media.

2.3 Interpretability

Several works explored interpretability using attention towards high-stakes decision making [4]. However, recent research sparked a debate on whether attention may provide what we consider as explanations which are faithful to the predictions of the model [20, 29]. From [18], we have seen that using other methods such as post-hoc interpretability do not provide us with explanations that may be seen to have high fidelity.

Towards this end, in this work we explore utilizing extracted entities which can be regarded as providing explanations to clinicians for gaining insights into Long COVID. We are the first, to the best of our knowledge, to present a novel approach for enhancing state-of-the-art Name Entity Recognition (NER) models with UMLS Metathesaurus and a benchmark social media dataset to improve entity-extraction related to Long COVID from social media.

3 Dataset

In this section, we explain data set acquisition, de-identification, preparation and filtering Long COVID self reports details.

3.1 Data Acquisition and De-identification

The dataset was acquired using academic access to Twitter API v2. Using the API, we acquired $466,651$ data points between August 2019 and June 2021. Each data point contains a tweet, timestamp, geographical coordinates if available, user's location (user-defined) and user's profile description. All tweets were retrieved using a list of hashtags (e.g., `longcovid`, `covidlong`, `longhauler`, etc.) and words included in tweets (e.g., `long-hauler`, `chronic symptoms`, `long-term effects`) among others. Some exclusion words related to collaborators' products were also part of the search criteria. Retweets, replies, quotes and nullcast were also excluded from the dataset, as well as any tweets that were not in English. Moreover, apart from the inclusion and exclusion criteria, all usernames and mentions were hashed to pseudonymize the data. Following the de-identification step, all URLs and special characters were removed. We also transformed all text to lowercase, and expanded contractions.

The average character length in our clean dataset is 133, and average word count is 23. Top 5 unigrams are `long`, `haul`, `covid`, `term`, and `effects`. Top 5 bigrams are `long haul`, `long term`, `term effects`, `long covid`, and `covid 19` Top 5 trigrams are `the long haul`, `for the long`, `long term effects`, `the long term`, and `a long haul`.

3.2 Filtering Long COVID Self Reports

We utilize ReGEX filters using personal pronouns (e.g. *I*, *Me*, *My*, *Mine*, *Myself*) and expressions of feeling (e.g. *feel*, *experience*, *symptoms*) to capture self reports and exclude any tweets from news outlets or other irrelevant discussions. We were able to extract self reports with 78% accuracy on a manually curated dataset of 1000 identified tweets using this simple regex based filtering method. Although this method cannot capture all self-reports, it is effective enough to collect a large number of relevant tweets for our analysis. As future work, the method to identify self reports can be replaced by a text classifier given enough annotation data. Such a method could both improve the filtering accuracy and also reduce the irrelevant tweets.

4 Methodology

We made use of the following transformer-based models to extract entities from the tweets by utilizing the n2c2 (2010) dataset [26] for fine-tuning:

- **COVID-Twitter BERT v2**[14]: A BERT-large uncased model pretrained on 97M unique tweets. We utilize this model in our experiments as several COVID-related terms are contained within the training corpus which may serve as a strong base model for subsequent fine-tuning on entity-extraction tasks.
- **UmlsBERT**[10]: This model considers the connection between medical works through the usage of Concept Unique Identifiers (CUI) of the Unified Medical Language System (UMLS) software that brings together a number of health and biomedical vocabularies to support interoperability[3].

In addition, we passed our corpus of tweets through UMLS' MetaMapLite [6] to augment entities extracted by the transformer models as shown in Figure 2. MetaMapLite is a tool that uses NLP and computational linguistic techniques to extract entities associating with UMLS' Concept Unique Identifiers (CUIs). CUIs capture a wide range of clinical concepts that fall under various categories, or semantic types, such as *signs and symptoms*, or *sosy* (e.g., *coughing*), and *disease or syndrome*, or *dsyn* (e.g., *Influenza*).

To compensate for MetaMapLite's limited coverage of colloquial expressions (e.g., *brain fog*), we introduced an additional approach, also shown in Figure 2. We started with the AMIA Task 3 dataset, which consists of clinical concepts from tweets (e.g., symptoms, adverse drug reactions) and their corresponding, human-assigned Medical Dictionary for Regulatory Activities (MedDRA) codes [19]. MedDRA is a medical terminology dictionary that is frequently used by regulatory authorities and the biopharmaceutical industry. Using the UMLS Metathesaurus, a biomedical thesaurus that links synonymous names from over 200 source vocabularies, we first mapped these MedDRA codes to CUIs. Then, for each concept in the AMIA dataset, we searched our tweets for matches wherein, i) the AMIA concept appears in the text and, ii) it was not already captured by MetaMapLite.

For N tweets, we have extracted entities from human analysis given by $\{e_1, e_2, ...e_n\}\forall e_i \in E$ and model analysis given by $\{e'_1, e'_2, ...e'_n\}\forall e'_i \in E'$. We calculate $MatchCount$ and $MatchCount'$ as defined by Equation 1 and Equation 2, where E and E' are the set of entities extracted by humans and the model, respectively, with duplicated extracted entities removed. Results of 200 tweets provided to annotators from this process are shown in Table 3, having trained 4 annotators on this task of entity extraction.

$$MatchCount = \frac{\sum_{n=1}^{N} E}{\sum_{n=1}^{N} (E \cup E')} \tag{1}$$

$$MatchCount' = \frac{\sum_{n=1}^{N} E'}{\sum_{n=1}^{N} (E \cup E')} \tag{2}$$

Tweet	Human Evaluation	Model Output	Metamap + AMIA Output
Still recovering from #Covid, but fatigue and pots are persistent - I feel my heart rate going crazy. Tried resting, antivirals and some home remedies. Next week I'll get my second vaccine. #longcovid, wish me luck!	#Covid, fatigue,pots, heart rate, resting, antivirals, vaccine,#longcovid	covid, fatigue, pots, my heart rate, resting, antivirals, home remedies, my second vaccine, longcovid	pots, fatigue, heart rate, crazy
I cannot taste or smell my food, my energy levels are low, I can barely sleep and getting out of bed is hard. COVID sucks! #longhauler	cannot taste or smell, low energy levels,barely sleep,getting out of bed is hard, COVID, #longhauler	my energy levels	bed, low, energy levels, sleep, taste
I got covid 3 months ago. Hard to concentrate the same way as before getting this virus... I'm having constant headaches and dealing with brain fog is exhausting. Anybody with the same symptoms? #LongCovid	covid,Hard to concentrate,constant headaches, brain fog,exhausting, #LongCovid	this virus, constant headaches, brain fog, the same symptoms	headaches, brain, exhausting, fog, brain fog
I am seeing an occupational therapist next week to get rid of my #longcovid symptoms. I hope I will climb the stairs without taking a breath halfway or using a steroid inhaler	occupational therapist,#longcovid, climb the stairs without taking a breath halfway, using a steroid inhaler	my # longcovid symptoms, a steroid inhaler	-

Table 1. Human Evaluation results are randomly sampled examples from tweets annotated by 4 trained annotators. Model Output results are observed by using a fine-tuned UmlsBERT model on the n2c2 (2010) dataset. The rightmost column represents strings that match MetaMapLite's extractions and terms in the AMIA Task 3 dataset.

Model	P	R	F1
UmlsBERT	**0.8742**	**0.8947**	**0.8843**
CT-BERT	0.7014	0.7375	0.7190

Table 2. Results of transformer-based models after fine-tuning on the n2c2 (2010) dataset for entity recognition.

5 Results and Discussion

From Table 1, we observe that transformer-based models have picked up most of the terms as humans. Occasionally, these models have extracted non-clinical

Annotator	$MatchCount$	$MatchCount'$
1	0.4529	0.5823
2	0.6206	0.6400
3	0.4228	0.4176
4	0.5760	0.4440

Table 3. Result of exact matches of human evaluation on 200 tweets identified by trained annotators with the model output. Annotator #3 and #4 have medical backgrounds.

descriptions of symptoms, e.g., `my heart rate` instead of `heart rate`. Other times, models failed to retrieve relevant terms, in example tweet 2 the model did not identify `cannot taste or smell`. As observed in Table 2, state-of-the-art model architectures such as UmlsBERT solely trained on datasets with the ability to capture "Problem", "Treatment", or "Test", albeit achieving strong performance on a test set from the same distribution (Table 2), are unable to generalize well to unseen samples, as is observed from the "Model Output" in Table 1. Furthermore, model architectures such as CT-BERT show poor performance due to the domain shift between the pre-training and fine-tuning datasets.

The terms extracted by MetaMapLite, along with terms in the AMIA Task 3, provide a promising dataset for fine-tuning the transformer models. We hypothesize that this could produce a model that better captures the idiosyncrasies of social media text; however, in a preliminary experiment, the model performed poorly when trained on only MetaMapLite results of the "sosy" semantic type, combined with AMIA Task 3 data. This is likely a reflection of i) sparse labelling —MetaMapLite often captures fewer entities than are actually present, and ii) incorrect labels, as there are frequent examples of non-clinical terms that are captured. To improve performance, we would like to both expand the dataset for fine-tuning and improve its quality. This would benefit greatly from the input of clinical SMEs capable of informing the inclusion of additional semantic types, and guidelines on limiting inaccurate or erroneous MetaMapLite results.

Conclusion and Future Work

In this work, we propose utilizing entity extraction methods as a means to provide insights into patient self-reported experiences with post-acute sequelae of SARS-CoV-2 (PASC), or Long COVID. We evaluate the performance of —(1) state-of-the-art transformer-based models fine-tuned on the n2c2 (2010) dataset and (2) data augmentation using entities extracted by MetaMapLite, alongside terms from the AMIA Task 3 dataset. We observe that, although this generally produces sensible results, it still falls short of human assessments of clinical entities, and frequently misses key terms while embellishing others with superfluous text. Future work would further explore the use of MetaMapLite, and datasets such as the one from AMIA Task 3, to annotate a dataset of tweets for fine-tuning the NER models while capturing the idiosyncrasies of social media text.

Ethical Considerations

A number of considerations are important, for example, if the present work can be extended more broadly to monitor patient discussions on their experience with other illnesses. Obviously, the first challenge is whether the use of such data meets the ethical requirements [30]. In our study, we obtained an ethics opinion suggesting de-identified publicly available data can be analyzed for generating trends and insights if no individual data are shared and risk of re-identification is extremely low. Based on this opinion, we implemented an anonymization step early in the dataset development process and restricted updating of the dataset made available for analyses. We see future studies will have to carve their paths to secure ethics approval and implement the required actions. This aligns with the recommendations from Chiauzzi and Wicks [5] that similar data science projects need to assure the participants in their studies who have not formally consented that their anonymity will not be compromised, or risk of harmful outcomes is rare. Nevertheless, Staccani and Lau highlight that patient acceptance of social media use for clinical trial surveillance could be favorable [22]. Their findings underline the need to seek social media users' opinion and perception on use of their social media posts for studying disease characteristics and surveillance purposes in public health.

Another important consideration is applying these approaches to other conditions. We understand that this may be an ambitious recommendation given that there is sparse labelling in the social media datasets. We overcame this challenge by casting a wide net of subject matter experts from academia, industry, and from expertise spanning clinical medicine, clinical research, epidemiology, and pharmaceutical research. Combined, these individuals provided insights about data curation strategies including but not limited to, selection of filters, data pre-processing, labelling of limited data, and scaling of this labelling to large datasets. These experts also participated in evaluating the quality of labelled data. Given our experience, involving a multidisciplinary group to expand these approaches to other conditions is feasible given that patients are expressing their views on social media (e.g., another emerging infection, high incidence cancer, unknown illness). Moreover, we suggest involving a group of patients to include patient reported outcomes in real-time and further inform the key word search strategy.

Acknowledgement

The authors would like to thank Vector Institute for making this collaboration possible and providing academic infrastructure and computing support during all phases of this work. We would also like to thank Antoaneta Vladimirova and

Celine Leng from Roche, Esmat Sahak from University of Toronto for their support throughout this project, as well as Dr. Angela Cheung from the University Health Network for her expertise and guidance.

References

1. Aucott, J.N., Rebman, A.W.: Long-haul covid: heed the lessons from other infection-triggered illnesses. The Lancet **397**(10278), 967–968 (2021)
2. Bhambhoria, R., et al.: A smart system to generate and validate question answer pairs for covid-19 literature. In: Proceedings of the First Workshop on Scholarly Document Processing. pp. 20–30 (2020)
3. Bodenreider, O.: The unified medical language system (umls): integrating biomedical terminology. Nucleic acids research **32 (Database issue)**, 267–270 (2004), https://doi.org/10.1093/nar/gkh061
4. C., E., et al.: RETAIN: interpretable predictive model in healthcare using reverse time attention mechanism. CoRR **abs/1608.05745** (2016), http://arxiv.org/abs/1608.05745
5. Chiauzzi, E., Wicks, P.: Digital trespass: ethical and terms-of-use violations by researchers accessing data from an online patient community. Journal of Medical Internet Research **21**(2), e11985 (2019)
6. Demner-Fushman, D., et al.: MetaMap Lite: an evaluation of a new Java implementation of MetaMap. Journal of the American Medical Informatics Association **24**(4), 841–844 (01 2017). https://doi.org/10.1093/jamia/ocw177, https://doi.org/10.1093/jamia/ocw177
7. Domingo, F., et al.: Prevalence of long-term effects in individuals diagnosed with covid-19: a living systematic review. medRxiv (2021)
8. Guo, J., et al.: Mining twitter to explore the emergence of covid-19 symptoms. Public Health Nursing **37**(6), 934–940 (2020)
9. Jelodar, H., et al.: Deep sentiment classification and topic discovery on novel coronavirus or covid-19 online discussions: Nlp using lstm recurrent neural network approach. IEEE Journal of Biomedical and Health Informatics **24**(10), 2733–2742 (2020)
10. M., G., et al.: Umlsbert: Clinical domain knowledge augmentation of contextual embeddings using the unified medical language system metathesaurus (2020)
11. Mackey, T., et al.: Machine learning to detect self-reporting of symptoms, testing access, and recovery associated with covid-19 on twitter: retrospective big data infoveillance study. JMIR public health and surveillance **6**(2), e19509 (2020)
12. Magge, A., Klein, A., Miranda-Escalada, A., Al-garadi, M.A., Alimova, I., Miftahutdinov, Z., Farre-Maduell, E., Lopez, S.L., Flores, I., O'Connor, K., Weissenbacher, D., Tutubalina, E., Sarker, A., Banda, J.M., Krallinger, M., Gonzalez-Hernandez, G. (eds.): Proceedings of the Sixth Social Media Mining for Health (#SMM4H) Workshop and Shared Task. Association for Computational Linguistics, Mexico City, Mexico (Jun 2021), https://aclanthology.org/2021.smm4h-1.0
13. Morgan, M., et al.: Information extraction for social media. In: Proceedings of the Third Workshop on Semantic Web and Information Extraction. pp. 9–16 (2014)
14. Müller, M., et al.: Covid-twitter-bert: A natural language processing model to analyse covid-19 content on twitter. arXiv preprint arXiv:2005.07503 (2020)
15. Phillips, N.: The coronavirus is here to stay-here's what that means. Nature **590**(7846), 382–384 (2021)

16. Pradhan, S., et al.: Semeval-2014 task 7: Analysis of clinical text. In: Proc. of the 8th International Workshop on Semantic Evaluation (SemEval 2014. Citeseer (2014)
17. Qin, L., et al.: Prediction of number of cases of 2019 novel coronavirus (covid-19) using social media search index. International journal of environmental research and public health **17**(7), 2365 (2020)
18. Rudin, C.: Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead (2019)
19. Sarker, A., Gonzalez-Hernandez, G.: Overview of the second social media mining for health (smm4h) shared tasks at amia 2017. In: Proc. of the 2nd Social Media Mining for Health Research and Applications Workshop co-located with the American Medical Informatics Association Annual Symposium (AMIA 2017) (2017), http://ceur-ws.org/Vol-1996/
20. Sha, Y., Wang, M.D.: Interpretable predictions of clinical outcomes with an attention-based recurrent neural network. In: Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics. pp. 233–240 (2017)
21. Smailhodzic, E., et al.: Social media use in healthcare: A systematic review of effects on patients and on their relationship with healthcare professionals. BMC health services research **16**(1), 1–14 (2016)
22. Staccini, P., Lau, A.Y., et al.: Social media, research, and ethics: Does participant willingness matter? Yearbook of Medical Informatics **29**(01), 176–183 (2020)
23. Sudre, C., et al.: Attributes and predictors of long covid. Nature medicine **27**(4), 626–631 (2021)
24. Suominen, H., et al.: Overview of the share/clef ehealth evaluation lab 2013. In: International Conference of the Cross-Language Evaluation Forum for European Languages. pp. 212–231. Springer (2013)
25. Tsao, S., et al.: What social media told us in the time of covid-19:a scoping review. The Lancet Digital Health (2021)
26. Uzuner, Ö., et al.: 2010 i2b2/va challenge on concepts, assertions, and relations in clinical text. Journal of the American Medical Informatics Association **18**(5), 552–556 (2011)
27. Vijayan, T., et al.: Trusting evidence over anecdote: Clinical decision making in the era of covid-19. BMJ (July 2020), https://blogs.bmj.com/bmj/2020/07/23/trusting-evidence-over-anecdote-clinical-decision-making-in-the-era-of-covid-19/
28. Wang, Y., et al.: Examining risk and crisis communications of government agencies and stakeholders during early-stages of covid-19 on twitter. Computers in human behavior **114**, 106568 (2021)
29. Wiegreffe, S., Pinter, Y.: Attention is not not explanation. CoRR **abs/1908.04626** (2019), http://arxiv.org/abs/1908.04626
30. Williams, M.L., Burnap, P., Sloan, L.: Towards an ethical framework for publishing twitter data in social research: Taking into account users' views, online context and algorithmic estimation. Sociology **51**(6), 1149–1168 (2017)