An Interactive Analysis of User-reported Long COVID Symptoms using Twitter Data

Lin Miao[1], Mark Last[1], Marina Litvak[2]
Ben-Gurion University of the Negev[1], Shamoon College of Engineering[2]
P.O.B. 653 Beer-Sheva 8410501 Israel[1], 56 Bialik St. Beer-Sheva 8410802 Israel[2]
miaol@post.bgu.ac.il, mlast@bgu.ac.il, litvak.marina@gmail.com

Abstract

With millions of documented recoveries from COVID-19 worldwide, various long-term sequelae have been observed in a large group of survivors. This paper is aimed at systematically analyzing user-generated conversations on Twitter that are related to long-term COVID symptoms for a better understanding of the Long COVID health consequences. Using an interactive information extraction tool built especially for this purpose, we extracted key information from the relevant tweets and analyzed the user-reported Long COVID symptoms with respect to their demographic and geographical characteristics. The results of our analysis are expected to improve the public awareness on long-term COVID-19 sequelae and provide important insights to public health authorities.

1 Introduction

The COVID-19 pandemic has affected millions of people all over the world. Despite the growing knowledge of COVID-19, much still remains unclear, especially potential long-term health consequences.

The term of Long COVID was brought up by the patients on Twitter in May 2020, in order to express their long-term COVID illness (Callard and Perego, 2021). Many Long COVID sufferers shared their persistent symptoms on social media bringing numerous discussions of similar symptoms experienced by others. Long COVID, also known as post COVID-19 syndrome, has no strict definition. The CDC of the United States[1] describes Long COVID as symptoms for four or more weeks after the infection, however, WHO[2] and British National Institute for Health and Care Excellence (NICE)[3] suggest three months after onset of COVID-19.

Though the majority of infected people experience mild symptoms with no necessity of hospitalization, the post-COVID syndrome is being reported by not just hospitalized patients. Therefore, massive user-generated Long COVID data available on social media has a significant value for tracking and analyzing the long-term syndrome.

Thus, in this work, we aim to apply Natural Language Processing (NLP) approaches to explore the characteristics of Long COVID symptoms reported by the Twitter users in terms of the patient gender, age, and location, as well as in terms of the symptoms duration. By extracting and analyzing key information from Long COVID-related tweets, we can discover less known chronic physical or mental conditions experienced by large groups of COVID-19 patients, and explore the relations between symptoms and demographic or geographic characteristics of patients. Moreover, we also seek to study the Long COVID evolution over time. To address this need, we compare the results of datasets collected in different time periods.

As part of this study, we developed an online dashboard[4] to visualize the analysis of Long COVID symptoms harvested from Twitter. A snapshot is shown in Fig. 1. This interactive dashboard provides multi-scale information and insights.

Our contributions can be summarized as follows:

- We build and publish two repositories of Long COVID-related tweets, which include user-generated reports on Long COVID experience from different periods of time[5].
- We conduct a comprehensive analysis of the

[1]https://www.cdc.gov/coronavirus/2019-ncov/long-term-effects/index.html?CDC_AA_refVal=https%3A%2F%2Fwww.%cdc.gov%2Fcoronavirus%2F2019-ncov%2Flong-term-effects.html

[2]https://www.who.int/publications/i/item/WHO-2019-nCoV-Post_COVID-19_condition-Clinical_case_definition-2021.1

[3]https://www.nice.org.uk/news/article/nice-rcgp-and-sign-publish-guideline-on-managing-the-long-term-effects-of-covid-19

[4]https://longcovid-dashboard.herokuapp.com/

[5]https://github.com/Lin1202/Longcoivd/blob/main/ long-covid_tweets.tar

Proceedings of the 2nd Workshop on Deriving Insights from User-Generated Text, pages 10 - 19

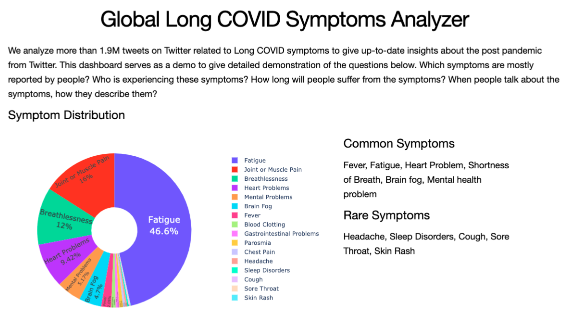


Figure 1: Snapshot of the dashboard

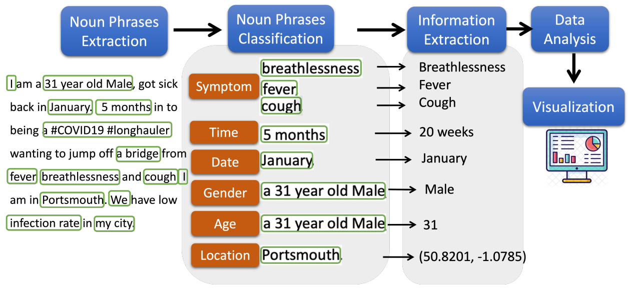


Figure 2: Methodology pipeline using synthetic data example

generated Long COVID datasets, which can provide insights to decision-makers and researchers.

- We explore several noun phrase classification models for information extraction from tweets to accurately recognize the long-term sequelae.
- We develop an online dashboard for interactive analysis of Long COVID symptoms from different perspectives.

2 Related Work

In response to the COVID-19 pandemic, extensive research has been conducted to help the healthcare community respond to this unprecedented emergency. As the concerns of COVID-19 long-term consequences are rising, more efforts are being invested in this topic. Current research about Long COVID uses standardized questionnaires or medical assessment to follow up the long-term symptoms of patients with clinical records (Carfì et al., 2020; Blomberg et al., 2021). Due to the lack of sufficient data about long-term COVID-19 complications, some studies explore the COVID-19 sequelae through the review of earlier papers (Mitrani et al., 2020; Kumar et al., 2021; Willi et al., 2021).

Numerous studies are using NLP approaches to contribute to the global response to this pandemic crisis. For instance, Silverman et al. (2021) use NLP pipeline to extract COVID symptoms from unstructured notes. With the use of NLP algorithms, Cury et al. (2021) assess the CT imaging reports for tracking of COVID-19 pandemic in the United States.

Data from social media is widely used for COVID symptom analysis (Sarker et al., 2020; Krittanawong et al., 2020). However, very few studies focus on Long COVID with social media attributes. Singh and Reddy (2020) and Banda et al. (2020) analyzed the long-term symptoms distribution by mining and manually reviewing tweets of around 100 self-reporting users. Sarker and Ge (2021) analyzed the major Long COVID symptoms distribution by extracting symptoms from posts on Reddit, using approximate matching approach based on an expanded meta-lexicon. They mainly analyzed the major symptoms distribution.

In this work, we utilize various NLP approaches to explore the Long COVID symptoms reported on Twitter. We aim to conduct multi-scale analysis of Long COVID symptoms, including not only the symptoms distribution and duration but also the effect of demographic and geographical patient characteristics.

3 Methodology

Our data analysis pipeline is demonstrated in Fig. 2, using a synthetic tweet based on several real tweets as an example. First, noun phrases (NPs) are identified in each tweet and then classified to different categories. Next, Long COVID-related information is extracted from the identified NPs for further analysis.

3.1 Noun Phrase Classification

We extract relevant information from tweets by identifying NPs from seven categories, as shown in Table 1. As observed, some NPs may carry more than one information category. For example, "my 31 year old daughter" contains 'age', and 'gender.' As such, we regard the NPs classification as a multi-label and multi-class classification task. After manually labeling some data, we aim to train a supervised classification model. In this work, we evaluate and compare the following NP classification models: (a) Support vector machine (SVM);

(b) Bidirectional Gated Recurrent Unit (GRU); (c) Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2019). The most accurate model is selected for this stage.

3.2 Information Extraction

After identifying NPs implicating valuable information, we process and extract information from these NPs for further analysis.

Symptom Categorization At this stage, we try to identify symptom-related NPs. The symptom classification task needs to capture the synonyms for each symptom category. Because text on Twitter is informal, the symptoms are not named in a consistent or complete way. We chose to overcome this problem using BERT embeddings. BERT is not only capable of providing similar embeddings for close meanings, it also gives contextualised embeddings. Additionally, as a Masked Language Model, BERT produces significant results in understanding an incomplete text. Therefore, we built the classification model by fine-tuning BERT to identify all the symptom-related NPs.

Gender Extraction We label NPs as "Gender" if they include gender information about the reported person, such as, "my daughter", "my husband". Also, some self-reported tweets mention the gender explicitly. We extract the patient gender ("male" or "female") from the relevant NPs by conducting binary classification.

Symptom Duration Calculation Considering the different definitions for Long COVID, we chose to use 'week' to measure the duration of the symptoms. It is observed that the users may report the time period of their Long COVID symptoms or the date of diagnosis, both of which can provide the duration of the symptoms. Therefore, we extract the symptom duration from NPs labeled as "Time" or "Date". The NPs classified as "Time", which are representing the time period, are converted into weeks directly. As for the NPs of "Date" category, which are expressing the date of the reported diagnosis, the duration is calculated by subtracting the creation date of the tweets.

Age Extraction The user age in years is extracted from the NPs assigned to the "Age" category by converting number text to integers and extracting the integers. Any extracted values that are out of age range ([0,100]) are then filtered out.

Location Process The geolocation information considers "Location" NPs and the reported locations in account profiles. However, if the geolocation information retrieved from the NP and the account profile are nonidentical for a single tweet, we keep the geolocation information from the NP. Later, the location information is converted into coordinates.

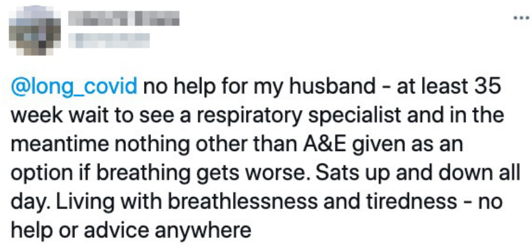
3.3 Data Analysis

In this work, we conduct analysis of three categories: demographic analysis, geographical analysis, and textual content analysis.

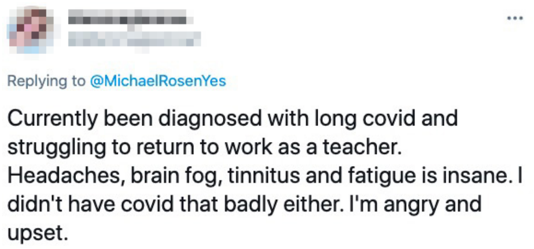
Demographic Analysis We analyze the distribution of the reported symptoms. Aligning the creation date of the tweets and the mentioned symptoms, we explore how the symptom distributions evolve through the timeline. To explore the associations of other features with symptoms, we associate the extracted information with the symptoms mentioned in the same tweet. Associated with gender information, we analyze the gender distribution of symptoms, to compare the symptoms experienced by men and women. Associated with gender and age/duration, we also present the joint distributions of gender, age/duration and several major symptoms respectively. The average age and duration of each symptom are calculated and demonstrated.

Geographical Analysis We visualize the locations on a global map, marked for different symptoms. The distribution of each symptom can be clearly seen on the map, providing a geographical perspective.

Textual Content Analysis To drill down into the tweets content, we generate a word cloud for the data of each month. The word cloud presents the most frequent words related to symptoms. Word cloud could enable to discover new symptoms and the interaction of several different symptoms. We use word distance with symptom words to filter out the frequent words that are irrelevant to Long COVID. Each word is represented by a word2vec vector. For each word, we calculate the distance to each vector in the symptom list, then keep the closest distance as the score for this word. Later, we sort these words by their scores from closest to farthest. The top 100 words are used for generating the word cloud.



(a) Example 1



(b) Example 2

Figure 3: Examples of Long COVID-related tweets

4 Case Study

Based on the proposed pipeline, we conducted in-depth analysis of Long COVID symptoms reported on Twitter using a sample of Long COVID-related tweets between May to December 2020, and a sample of October 2021. To explore the Long COVID evolution over time, we compared the results of May-December 2020 and October 2021.

4.1 Data

Leveraging Twitter's streaming API, Chen et al. (2020) are using keywords to continuously collect a significant amount of COVID-19 tweets and periodically release it for research use. We used specific keywords (shown in Appendix A.1) to retrieve relevant tweets about Long COVID symptoms from this public coronavirus twitter dataset. We built two datasets: one is containing Long COVID-related tweets from 1st May to 31st December 2020; the other one is containing Long COVID-related tweets of October 2021. After removing non-English tweets and duplicates, our two datasets contain 2.3M relevant tweets, in which the amount is approximately 3.2% of the original dataset for the same period of time. Some Long COVID tweet examples are shown Fig. 3. Using keyword-based approach to gather tweets might exclude some relevant ones because of missing certain keywords or misspellings (as tweets are notoriously noisy). To estimate the False Negative (ratio of missing tweets) in our collection, we manually verified 1432 tweets randomly selected from the original COVID-19 dataset. As a result, we got True Positive Rate of 78.8%, finding only two Long COVID-related tweets, which were missing in our collection. Hence, we may assume that our keyword-based approach with only 0.1% False Negative Rate is sufficient to cover most relevant tweets from the source dataset.

Category	Description	#Labeled
Symptom	symptoms reported in tweets	340
Time	time period mentioned in tweets	179
Date	date mentioned in tweets	68
Age	age reported in tweets	44
Gender	gender reported in tweets	30
Location	location reported in tweets	45
None	none of the above	436

Table 1: Noun phrase categories and labeled data for NPs classification

4.2 Noun Phrases Extraction and Classification

We used spaCy[6] for extracting the NPs from each tweet.

For NPs classification, one of the authors manually labeled a small subset of data from the 2020 dataset for training a supervised classification model. The details of the labeled data are shown in Table 1. The labeled data was randomly split into training and testing sets (80/20). Using the labeled data, we evaluated the following text representations and classification algorithms to select the most accurate model for NPs classification: (a) SVM using tf-idf vectors for NPs representation; (b) Bidirectional GRU using the Global Vectors for Word Representation (GloVe) model (Pennington et al., 2014) for NPs representation; (c) BERT, fine-tuned on our NPs classification task.

We applied TfidfVectorizer[7] for tf-idf text representation, utilized 200d GloVe for text representation of NPs, and implemented GRU in Keras[8]. The maximum sequence length was set to 5. We relied on the Sigmoid activation function and learned the weights using the Adam optimizer and Cross Entropy loss. We used a typical batch size of 32. We fine tuned BERT using the pre-trained English bert-base-cased model(Devlin et al., 2019), which has 12 transformer layers, 12 self-attention heads, and a hidden size of 768. We applied a pre-trained BERT

[6]https://spacy.io/
[7]https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text. TfidfVectorizer.html
[8]https://keras.io/

Classifier	Accuracy	HammingLoss
SVM_tf-idf	0.72	0.043
GRU_GloVe	0.76	0.049
BERT	0.89	0.022

Table 2: Results of different classifiers for NPs classification

Category	Precision	Recall	F1
Age	1.00	1.00	1.00
Date	0.92	0.92	0.92
Gender	1.00	0.67	0.80
Location	1.00	1.00	1.00
None	0.90	0.88	0.89
Symptom	0.94	0.88	0.91
Time	1.00	1.00	1.00
Micro avg	0.93	0.89	0.91
Macro avg	0.97	0.91	0.92
Weighted avg	0.93	0.89	0.91
Samples avg	0.90	0.89	0.90
Accuracy	0.89		
Hamming Loss	0.022		

Table 3: The detailed results of NPs classification using fine-tuned BERT

Symptom	#Labeled	P	R	F1
Blood Clotting	24	1.00	1.00	1.00
Brain Fog	76	0.96	1.00	0.98
Breathlessness	186	0.96	0.98	0.97
Chest Pain	45	0.95	0.95	0.95
Cough	16	1.00	1.00	1.00
Fatigue	201	0.95	1.00	0.98
Fever	45	1.00	0.80	0.89
Gastrointestinal Problems	30	1.00	1.00	1.00
Headache	32	0.77	1.00	0.87
Heart Problems	138	0.97	1.00	0.99
Joint or Muscle Pain	147	0.93	0.93	0.93
Mental Problems	69	0.92	1.00	0.96
Parosmia	18	1.00	1.00	1.00
Skin Rash	18	1.00	0.83	0.91
Sleep Disorders	24	1.00	1.00	1.00
Sore Throat	21	0.80	0.57	0.67
None	31	1.00	0.29	0.44
Macro avg		0.95	0.90	0.91
Weighted avg		0.95	0.95	0.94
Accuracy		0.95		

Table 4: The detailed results of symptom classification using fine-tuned BERT

model to produce dense vector representations for NPs, attaching a dense layer and a softmax layer to fine tune the model for our task. We used the Adam optimizer and Cross Entropy loss for training, with 1e-5 as the initial learning rate, and batch size 32.

The results of different NP classifiers are shown in Table 2. The fine-tuned BERT classifier outperformed others, so we apply it for the NPs classification. More detailed results of fine-tuned BERT are shown in Table 3.

Since we are only interested in the tweets containing Long COVID symptoms, the tweets without "Symptom" NPs were also filtered out. As a result, 61% of the data was retained for further steps.

4.3 Information Extraction from Noun Phrases

Referring to Long COVID Wikipedia[9] and other related studies (Del Rio et al., 2020; Olliaro, 2021), we summarized the long-term symptoms to 16 classes (shown in Table 4). We used semi-automatic method to label a 1K sample of symptom NPs from the 2020 dataset for training supervised classification models. First, we applied K-means with K=17, for dividing all NPs to 17 clusters (for 16 symptom classes and 1 for miscellaneous NPs describing symptoms that are not included in our list). Then, we manually corrected the clustering results, where 32.6% of each cluster instances were corrected on average. Subsequently, we used this labeled data to fine-tune BERT for a symptom NPs classification task. The labeled data was randomly split into 80% for training and 20% for test. Details of the model performance are shown in Table 4. Given this model, all of the symptom NPs were automatically labeled. We used the same configurations, as shown in 4.2 for the NPs classification, for fine tuning BERT for symptoms categorization.

For gender NPs classification, we used the zero-shot text classification pipeline[10], which is based on the Bart (Lewis et al., 2020) model pretrained on Multi-Genre Natural Language Inference (MultiNLI) corpus[11].

We applied geopy[12] to convert the location information into geographic coordinates.

When generating word clouds for the symptom-related tweets, we calculated semantic similarity between each word and any of explored symptoms using the cosine similarity between their 300d word2vec vectors. Words with low similarity to all symptoms were discarded.

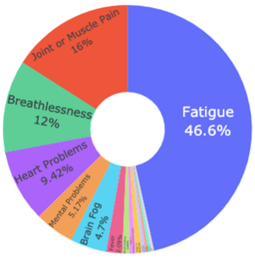
More detailed performance of information extraction is shown in Appendix A.2.

[9]https://en.wikipedia.org/wiki/Long_COVID

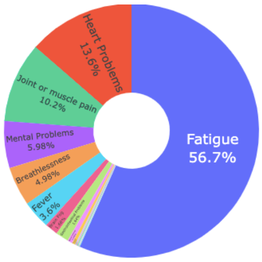
[10]https://discuss.huggingface.co/t/new-pipeline-for-zero-shot-text-classification/681

[11]https://cims.nyu.edu/ sbowman/multinli/

[12]https://pypi.org/project/geopy/



(a) May-December 2020



(b) October 2021

Figure 4: Distribution of symptoms

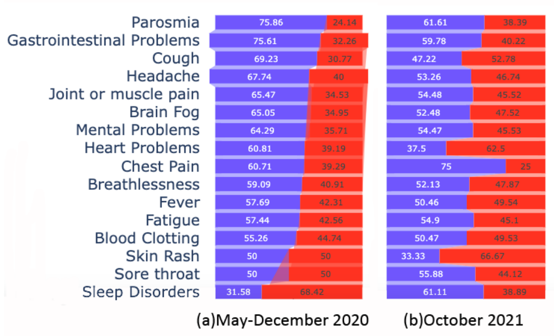


Figure 5: Gender distribution of each symptoms (Male: red; female: blue)

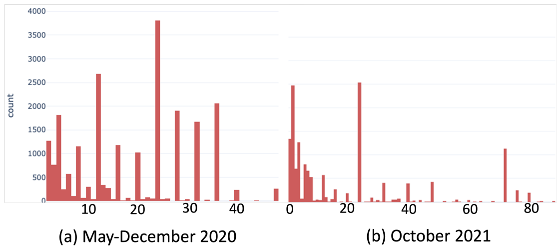


Figure 6: Duration(weeks) distribution

5 Results Analysis

After conducting information extraction with multiple pre-trained models, we performed comprehensive analysis of symptom-related data. We compared the results of May-December 2020 and the results of October 2021 to analyze the Long COVID evolution over time. The results of May-December 2020 were presented in an interactive dashboard.

Long COVID Symptoms Analysis Extracting the symptoms from the tweets, we analyzed the symptom distribution using the frequency of each symptom reported in each dataset. Fig. 4(a) shows the distribution of the symptoms in May-December 2020. From Fig. 4(a), we can see that the major symptoms are fatigue, breathlessness, joint or muscle pain, heart problems, and brain fog. Some rarely reported symptoms are headache, sleep disorders, cough, sore throat, skin rash.

Fig. 4 shows the comparison of symptom distribution in May-December 2020 and October 2021. Generally speaking, the top symptoms and rare symptoms stay approximately the same, except for the following changes. The percentages of fatigue and heart problems have increased from 46.6% to 56.7% and from 9.42% to 13.6%, respectively. The percentage of breathlessness and joint or muscle pain have decreased from 12% to 4.98% and from 16% to 10.2%, respectively. These changes may be explained by the mutations of the virus over time.

Symptoms Association with Gender We associated the extracted gender with the symptoms reported in the same tweet, analyzing the gender distribution with each symptom to explore the different experiences between men and women. Fig. 5 shows the gender distribution with symptoms of two datasets. Specifically, in our dataset of 2020, 62% of mentioned patients are reported as female, and 38% are reported as male. However, it is reported that about 30% Twitter users are female, and about 70% are male[13]. This may indicate that women are significantly more likely than men to suffer ongoing symptoms after COVID-19. Most of the symptoms appear to have an approximately equal distribution between men and women. Some very common symptoms, such as breathlessness, brain fog, and heart problems are more likely to happen to women than men. However, men are more likely to suffer from sleeping difficulties than women. And women are reported far more than men to suffer from parosmia and gastrointestinal problems.

Different from the 2020 dataset containing 62% female and 38% male, the dataset of October 2021 contains 54% female and 46% male. It still shows that women patients are more likely to be reported suffering Long COVID, however, the percentage of reported men patients is increasing. Similarly to the results of 2020, most symptoms show approximately equal distribution between men and women. The symptoms having big differences of distribution between men and women include cough and sleeping disorders.

[13]https://www.statista.com/statistics/828092/distribution-of-users-on-twitter-worldwide-gender/

Symptom	May-Dec 2020	Oct 2021
Blood Clotting	(17.898, 21.121)	(18.762, 27.016)
Brain Fog	(18.942, 20.195)	(17.156, 21.957)
Breathlessness	(17.510, 18.362)	(17.259, 20.020)
Chest Pain	(20.077, 22.510)	(8.373, 22.126)
Cough	(11.656, 14.386)	(7.177, 14.756)
Fatigue	(18.591, 19.074)	(19.788, 20.830)
Fever	(17.562, 19.207)	(14.743, 18.396)
Gastrointestinal Problems	(14.433, 16.872)	(14.858, 19.714)
Headache	(16.504, 19.029)	(10.129, 26.501)
Heart Problems	(18.692, 19.690)	(18.607, 20.303)
Joint or Muscle Pain	(18.543, 19.291)	(17.758, 20.041)
Mental Problems	(18.216, 19.515)	(22.178, 25.745)
Parosmia	(16.707, 18.956)	(15.263, 22.151)
Skin Rash	(10.241, 14.558)	(1.282, 35.117)
Sleep Disorders	(18.382, 23.617)	(10.141, 30.413)
Sore Throat	(16.472, 21.527)	(15.273, 30.998)

Table 5: The confidence intervals of symptom duration(weeks)

Symptom	May-Dec 2020	Oct 2021
Blood Clotting	(30.010, 36.076)	(28.692, 39.036)
Brain Fog	(35.513, 39.132)	(29.639, 35.804)
Breathlessness	(37.198, 38.987)	(31.097, 34.411)
Chest Pain	(35.191, 41.665)	(28.707, 60.403)
Cough	(16.184, 30.482)	(18.636, 46.613)
Fatigue	(35.641, 36.688)	(28.711, 29.681)
Fever	(26.829, 31.232)	(29.640, 33.714)
Gastrointestinal Problems	(38.418, 46.581)	(26.491, 32.196)
Headache	(35.162, 42.837)	(15.174, 30.825)
Heart Problems	(36.291, 38.292)	(28.497, 30.460)
Joint or Muscle Pain	(36.291, 38.292)	(30.348, 32.757)
Mental Problems	(35.916, 39.837)	(28.389, 31.763)
Parosmia	(36.460, 45.967)	(29.415, 42.908)
Sleep Disorders	(31.259, 44.740)	(8.391, 62.608)

Table 6: The age confidence intervals of each symptom

Symptoms Duration Associating symptoms with duration information, we analyzed for how much time people reported themselves to suffer from the Long COVID symptoms. Duration distributions of all symptoms of both datasets are shown in Fig. 6. In general, the duration of symptoms reported ranges from less than one month to more than ten months. From the comparison of symptoms duration distributions shown in Fig. 6, we can see that the distribution of October 2021 roughly follows the results of 2020. Most of the duration is less than half year, but the results of October 2021 contain a bigger proportion of duration, which is less than three months. It should be considered that some reported symptoms duration did not follow the same definition of Long COVID. However, considering the data of October 2021 observes longer time of COVID, some reported duration can reach 80 weeks.

Table. 5 shows the confidence intervals of symptoms duration in both datasets. In 2020 dataset, the duration for the most symptom categories is around five months. However, cough has been reported to last a relatively short time.

Symptoms Association with Age After extracting the age information, we linked the age and the symptom reported in the same tweet. Fig. 7 shows the distributions of age reported for 'long haulers' in our datasets and also compares the age distribution of all Twitter users[14]. The extracted age values in our datasets are mostly below 50. A possible reason is that people below 50 years old are more active on Twitter. However, despite the low percentages of age groups below 18 and above 50 among Twitter users, these age groups are reported more frequently to have Long COVID in our datasets. Notably, comparing the results of October 2021 to the results of 2020, the age group below 30 has a big proportion in the data of October 2021.

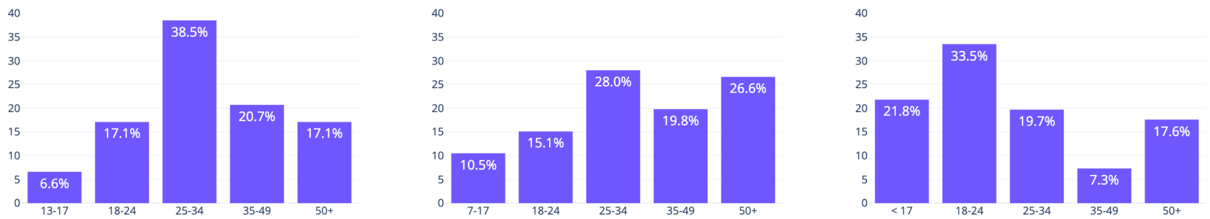
Table. 6 shows the age confidence intervals of each symptom. Comparing the results of October 2021 to the results of 2020, it can be observed that except for the increased age of cough and fever, the average age of patients experiencing other symptoms clearly decreased.

Symptoms Association with Geolocation Using the extracted geolocation information, we demonstrate geographic distributions of each symptom in the world map. As the COVID-19 pandemic has spread all over the world, the geographic distributions of May-December 2020 and October 2021 are approximately the same. Fig. 8 shows one example symptom "Joint Pain" association with geolocation based on the dataset of May-December 2020. Britain and USA have more twitter users reporting Long COVID symptoms than any other country. A possible explanation might be that people in Britain and USA are more aware about the long COVID symptoms. It can also be explained by these two countries being most popular on Twitter out of English-speaking countries[15].

Content Analysis In order to demonstrate the representative context of long COVID reports, we generated a word cloud for each month to present the most frequent symptom-related words. In gen-

[14] https://www.statista.com/statistics/283119/age-distribution-of-global-twitter-users/

[15] https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/



(a) Age distribution of Twitter users worldwide

(b) Age distribution in our dataset of May-December 2020

(c) Age distribution in our dataset of October 2021

Figure 7: Age distribution comparison of Twitter users and reported in our datasets

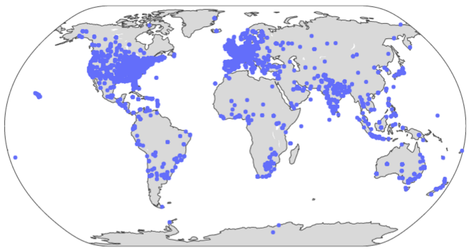
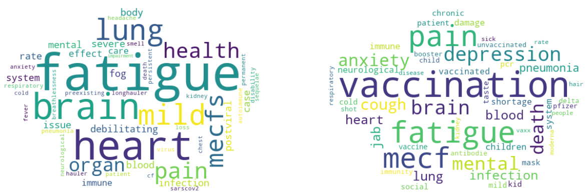


Figure 8: "Joint Pain" association with geolocation



(a) October 2020

(b) October 2021

Figure 9: Word clouds

eral, the most frequent words are related to the most frequent symptoms. By presenting the word cloud on monthly basis, we explored the dynamics of the frequent words. The examples of the word clouds are shown in Fig. 9. Comparing the word clouds of October 2020 and October 2021, we can see that some common symptoms are frequently discussed in both datasets. However, considering the vaccination progress in 2021, from Fig. 9(b) we can see that vaccination-related words are frequently discussed along with Long COVID symptoms. Notably, "depression", "anxiety", and "mental", which represent the symptoms of mental problems are more frequently discussed in the dataset of 2021.

In conclusion, the results of October 2021 are partially consistent with 2020. Notably, with the evolving COVID-19 pandemic including mutations and vaccination some characteristics of Long COVID symptoms appear to be evolving over time.

6 Discussion

Some important insights can be gained from the analysis. For example, women tend to experience Long COVID more frequently than men, which is similar to the findings of some medical studies (Sudre et al., 2021; Ortona and Malorni, 2021; Bai et al., 2021; Blomberg et al., 2021). Additionally, 'mental problems' is one of the top symptoms shown in our results, which is rarely mentioned as a common symptom in other works. In our work, 'mental problems' refers to depression, anxiety, loneliness and other mental or emotional health issues. Similarly, Sarker and Ge (2021) reported 55.2% users in the Reddit dataset experiencing mental problems since their COVID onset.

7 Conclusions

In this work, we conducted a comprehensive analysis of Long COVID symptoms reported by the Twitter users with respect to their demographic and geographical characteristics. The presented case study provides detailed information and important insights about multiple aspects of long-term symptoms. The comparative analysis of two periods of time (in 2020 and 2021) shows the consistent and the evolving characteristics of the Long COVID. Furthermore, an interactive online dashboard was built to visualize the results of the 2020 dataset. Limitations of this work include the possible effect of large amounts of noise in the Twitter data on our results. Besides, the data analyzed was limited to English tweets, which might not be representative of all segments of the world population affected by COVID-19. Thus, our future work will focus on analyzing larger amounts of data in multiple languages. Symptoms co-occurrence and presence of comorbidities may also be explored in the future work. Moreover, non-binary gender class may be taken into account in the future work as well.

Ethics Justification

This work is completely based on publicly accessible Twitter data. So there is no necessity for ethical approval. As the signing of the Twitter User Agreement by the users, no further user consent is required in this work for the data to be used. The results in this paper are not fully representative because the data is only from the Twitter social media platform and only English tweets are included.

Acknowledgements

This research was partially supported by an internal grant No. 87867211 from Ben-Gurion University of the Negev Corona Task Force.

References

Francesca Bai, Daniele Tomasoni, Camilla Falcinella, Diletta Barbanotti, Roberto Castoldi, Giovanni Mulè, Matteo Augello, Debora Mondatore, Marina Allegrini, Andrea Cona, et al. 2021. Female gender is associated with "long covid" syndrome: a prospective cohort study. *Clinical Microbiology and Infection*.

Juan M Banda, Gurdas Viguruji Singh, Osaid Alser, and Daniel Prieto-Alhambra. 2020. Long-term patient-reported symptoms of covid-19: an analysis of social media data. *medRxiv*.

Bjørn Blomberg, Kristin Greve-Isdahl Mohn, Karl Albert Brokstad, Fan Zhou, Dagrun Waag Linchausen, Bent-Are Hansen, Sarah Lartey, Therese Bredholt Onyango, Kanika Kuwelker, Marianne Sævik, et al. 2021. Long covid in a prospective cohort of home-isolated patients. *Nature Medicine*, pages 1–7.

Felicity Callard and Elisa Perego. 2021. How and why patients made long covid. *Social Science & Medicine*, 268:113426.

Angelo Carfì, Roberto Bernabei, Francesco Landi, et al. 2020. Persistent symptoms in patients after acute covid-19. *Jama*, 324(6):603–605.

Emily Chen, Kristina Lerman, and Emilio Ferrara. 2020. Tracking social media discourse about the covid-19 pandemic: Development of a public coronavirus twitter data set. *JMIR Public Health and Surveillance*, 6(2):e19273.

Ricardo C Cury, Istvan Megyeri, Tony Lindsey, Robson Macedo, Juan Batlle, Shwan Kim, Brian Baker, Robert Harris, and Reese H Clark. 2021. Natural language processing and machine learning for detection of respiratory illness by chest ct imaging and tracking of covid-19 pandemic in the united states. *Radiology: Cardiothoracic Imaging*, 3(1):e200596.

Carlos Del Rio, Lauren F Collins, and Preeti Malani. 2020. Long-term health consequences of covid-19. *Jama*, 324(17):1723–1724.

J. Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *NAACL-HLT*.

Chayakrit Krittanawong, Bharat Narasimhan, Hafeez Ul Hassan Virk, Harish Narasimhan, Zhen Wang, and WH Wilson Tang. 2020. Insights from twitter about novel covid-19 symptoms. *European Heart Journal-Digital Health*, 1(1):4–5.

Naresh Kumar, Dhiraj Wasnik, Harsh Vardhan, and MK Daga. 2021. Long term health sequelae of covid-19: A review. *Journal of Advanced Research in Medicine (E-ISSN: 2349-7181 & P-ISSN: 2394-7047)*, 8(1):9–18.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880.

Raul D Mitrani, Nitika Dabas, and Jeffrey J Goldberger. 2020. Covid-19 cardiac injury: Implications for long-term surveillance and outcomes in survivors. *Heart rhythm*, 17(11):1984–1990.

Piero L Olliaro. 2021. An integrated understanding of long-term sequelae after acute covid-19. *The Lancet Respiratory Medicine*.

Elena Ortona and Walter Malorni. 2021. Long covid: to investigate immunological mechanisms and sex/gender related aspects as fundamental steps for a tailored therapy.

Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.

Abeed Sarker and Yao Ge. 2021. Mining long-covid symptoms from reddit: characterizing post-covid syndrome from patient reports. *JAMIA open*, 4(3):ooab075.

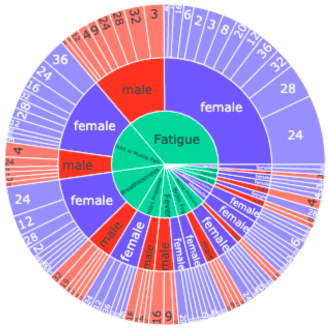
Abeed Sarker, Sahithi Lakamana, Whitney Hogg-Bremer, Angel Xie, Mohammed Ali Al-Garadi, and Yuan-Chi Yang. 2020. Self-reported covid-19 symptoms on twitter: an analysis and a research resource. *Journal of the American Medical Informatics Association*, 27(8):1310–1315.

Greg M Silverman, Himanshu S Sahoo, Nicholas E Ingraham, Monica Lupei, Michael A Puskarich, Michael Usher, James Dries, Raymond L Finzel, Eric Murray, John Sartori, et al. 2021. Nlp methods for extraction of symptoms from unstructured data for use in prognostic covid-19 analytic models. *Journal of Artificial Intelligence Research*, 72:429–474.

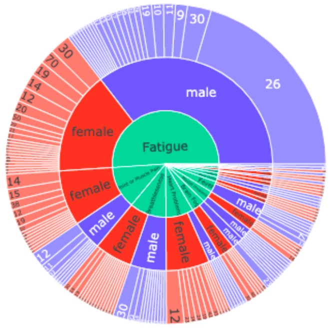
NPs Category	Detection Precision	Proportion
Symptom	95%	61%[1]
Patient Gender	98%	12%[2]
Patient Age	92%	5%[2]
Symptom Duration	98%	50%[2]
Patient Location	99%	23%[2]

Table A.1.1: Results of information extraction

[1] Proportion of symptom-related tweets in the dataset of May-December 2020.
[2] Proportion in symptom-related tweets.



(a) Joint distribution of symptoms with gender and duration



(b) Joint distribution of symptoms with gender and age

Figure A.3.1: Joint distribution in May-December 2020 Data

Shubh Mohan Singh and Chaitanya Reddy. 2020. An analysis of self-reported longcovid symptoms on twitter. *medRxiv*.

Carole H Sudre, Benjamin Murray, Thomas Varsavsky, Mark S Graham, Rose S Penfold, Ruth C Bowyer, Joan Capdevila Pujol, Kerstin Klaser, Michela Antonelli, Liane S Canas, et al. 2021. Attributes and predictors of long covid. *Nature medicine*, 27(4):626–631.

Sandra Willi, Renata Lüthold, Adam Hunt, Nadescha Viviane Hänggi, Donikë Sejdiu, Camila Scaff, Nicole Bender, Kaspar Staub, and Patricia Schlagenhauf. 2021. Covid-19 sequelae in adults aged less than 50 years: a systematic review. *Travel medicine and infectious disease*, page 101995.

A Appendix

A.1 Keywords list of searching tweets

Keywords related to Long COVID are (uncased): LongCovid, Long Covid, Long haul Covid, Long hauler, Chronic Covid Syndrome (CCS), Post Covid symptoms, Long lasting symptoms, Long-term symptoms, sequelae covid, persistent symptoms.

A.2 Performance of Information Extraction

To validate the automatic labeling results, we manually verified some labeled samples. The accuracy of the validation samples is 98%. More specifically, the errors are caused by gender bias of the pre-trained model. For example, " My kid" is labeled as "male", "My mother's friend" is labeled as "female". However, the accuracy shows that the approach is valid for gender extraction.

In order to investigate the accuracy of information extraction, we manually checked some sample sets for each class label after extraction. The results are shown in Table A.1.1. Besides, because our work aims to explore the association of symptoms with gender, age, duration, and geolocation, we are interested to see the distributions of these features in symptom-related tweets. We calculated the ratios of gender, age, duration, and location in symptom-related tweets separately. For duration extraction, NPs automatically labeled as "Time" and "Date" were utilized. About 50% of symptom-related tweets contain the relevant information, specifically 38% of them are labeled as "Time" and 12% as "Date". We summarized the key information detection accuracy and the distribution of the detected labels in Table A.1.1.

A.3 Joint Distribution

We linked symptoms with gender and duration getting the joint distribution, which demonstrates for how long time men and women were more likely to be reported experiencing certain symptoms.

We also present the joint distributions of age and gender in tweets reporting about long COVID symptoms, showing in which age group men or women were more likely to report experiencing certain symptoms. One example of the results of 2020 is shown in Fig. A.3.1, from which it can be seen that mostly women of 30 years old frequently reported to experience fatigue.