A Social Media Study on COVID-19 Long-Hauler Community: The Community That Is Redefining COVID-19

A, B, E

The COVID-19 pandemic has affected more than 189 million people worldwide so far. Many communities (such as minority communities) suffered disproportionately more difficulties throughout the pandemic. In this paper, we would like to focus one such community: COVID-19 long-haulers community. Long-hauler community consists of those people who get affected by Corona virus but their symptoms do not cure in couple of weeks; rather they experience lingering symptoms for months. The concerns of this community were initially ignored by health care providers primarily because of limited information. In this paper, we have analyzed the discussion of a private social media group dedicated to long-hauler community. Our analyses revealed the primary discussion topics of this community. It also showed how a minority community can stand by each other using social media groups at the time of crisis. We concluded the paper with long-term implications of our findings for health care systems and policies.

CCS Concepts: • **Computer systems organization** → **Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability.

Additional Key Words and Phrases: Long-hauler, social media, medical gaslighting, lingering symptoms

ACM Reference Format:
A. 2018. A Social Media Study on COVID-19 Long-Hauler Community: The Community That Is Redefining COVID-19. *J. ACM* 37, 4, Article 111 (August 2018), 22 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

The COVID-19 pandemic, a global health crisis, has caused deaths to more than 4 million people worldwide so far. In the USA only, more than 34 million people got infected by the virus and the numbers are still counting. The impact of this virus has been so severe that even trained healthcare professionals describe it as — "I have never seen anything like this before...ever" [2]. Since the pandemic has continued to spread for more than a year now, it has no longer remained a health crisis; rather it has become a human, economic, and social crisis. Many communities (such as minority communities) suffered a disproportionate amount of difficulties throughout the pandemic. In this paper, we would like to focus on one such community: COVID-19 long-haulers community.

Who belongs to the COVID-19 long-haulers community? Typically, mild or moderate COVID-19 symptoms last about two weeks for most people. But, some individuals experience lingering health problems even when they have recovered from the acute phase of the illness. In such patients, there is no longer a live coronavirus running amok in the body. If tested, the person would test negative for the coronavirus, but they might be severely debilitated nonetheless. People living with post-COVID syndrome for a long time are known as "COVID-19 long haulers", "long-COVID", or "post-acute sequelae of COVID-19" as referred by the National Institutes of Health (NIH). Long-hauler patients have faced a wide range of symptoms which made it extremely difficult for them to go back to their

Author's address: A, B, C, D, E, F.

0004-5411/2018/8-ART111 $15.00
https://doi.org/10.1145/1122445.1122456

usual routine. However, physical symptoms were not the only challenges that they faced during this pandemic. Long-hauler patients experienced a great deal of mental pain and anxiety since healthcare professionals initially did not believe this condition was a real one. In many cases, these patients were referred for psychiatric evaluations, although their conditions were not remotely related to their mental health.

The concept of downplaying physiological condition as a psychological problem such as stress, anxiety, or somatic symptom disorder (a mental condition when someone has excessive and unrealistic worries about their health) is not new. This concept is called "medical gaslighting". Women experience medical gaslighting the most because of the knowledge and trust gap of healthcare providers [37]. In the recent past, the concept received a lot of media attention because of the book titled "Doing Harm" [29] published by Maya Dusenbery where she shared her own traumatic experience of medical gaslighting, discussed more largely gender bias in medicine, and how it hurts women. Long-haulers have made this condition worse for both men and women.

Why were Long-hauler patients gaslit when so many initiatives were put to treat people infected by coronavirus? One primary reason was the lack of knowledge about this new virus and how different individuals might get infected by the virus. Even before much research has been done on this virus, healthcare providers started believing that COVID-19 would kill only a few (between 1% to 2% of the infected people) and the rest would only face "mild" symptoms for some time. Healthcare providers strictly followed the official description of COVID-19 symptoms and relied heavily on the COVID lab-test results. No one listened carefully to those people who were reporting unusual symptoms but tested negative for coronavirus. They were excluded from any possible help or valid answers only because either their conditions were not found "deadly" or they were not found COVID positive by standard tests. Already overwhelmed medical facilities with a huge number of critical patients did not have anyone left to address issues that Long-hauler patients were dealing with for months.

This ignorance motivated Long-hauler patients to seek help from each other. At the time of the pandemic, when social gatherings were almost impossible to arrange, these individuals formed Facebook groups to collaborate. They shared their experiences with other group members who were also experiencing similar symptoms or some unexplained symptoms that their healthcare providers could no address. In the last year, a substantial number of research projects have addressed various socio-economical challenges that people are still facing due to the pandemic. Unfortunately, to our knowledge, no prior work has focused their attention on long-hauler patients, their challenges, and initiatives. Because of growing interests and recent actions taken for addressing the conditions of Long-hauler patients, it is critical to understand how social media groups have helped them cope up with difficulties and challenges as a community when their complaints were mostly ignored by everyone else. One such group is the Facebook group called "COVID-19 Long Haulers Discussion Group".

This is a private Facebook group that was created on June 26th, 2020. The group is visible to everyone but requires approval from admins for joining. So far, it has 13,308 members and counting. The primary objective of this group is to emotionally assist Long-hauler patients. As mentioned in their information page," "we are ... here to offer experiences and support. You are not alone. This is not anxiety. Enter with kindness and grace." In the context of this long hauler social media community, we primarily asked the following research question:

What did the long-hauler community discuss in their private social media group? What can we learn from their discourse that may have long-term implications on long-hauler treatment facilities, rehabilitation programs, and most importantly public health care policies?

To address this research agenda, We manually collected all posts and comments posted in this group from November 3rd, 2020 till February 6th, 2021. Our dataset contained 186,860 entries which

included both posts and comments. We qualitatively annotate 600 of them to identify the main topics of discussion of this community. We also performed LDA, an unsupervised topic modeling algorithm, on the entire dataset to identify the naturally occurring topics of discussion in this group. Considering the outcomes of qualitative annotation and LDA topic modeling together, we identified two primary topics: 1) posts and comments regarding Long hauler symptoms and 2) posts and comments **not related to** Long hauler symptoms.

Next, we developed a group of machine learning classifiers that used features based on word embeddings, psycholinguistic attributes, open-vocabulary-based n-grams, and sentiments to identify symptom reporting posts and comments from the corpus. After demonstrating the best performing classifier to provide robust and stable performance with an AUC of 0.78, we machine label our entire dataset. We proceed our examination further on those posts and comments that were labeled as related to Long hauler symptoms. We identified all unique symptoms reported by the members of the community and designed a network-graph visualization using those symptoms. Each node of this visualization was one symptom and the symptoms that co-occurred were connected with an edge. The representation allowed us to situate all symptoms (reported by the members of the group) and their severity on a single plane, thereby producing a unique way to show the diversity and interconnection of the symptoms shared by the Long haulers community.

Finally, we identified the linguistic markers associated with the posts and comments of this Long-COVID community and built a vocabulary for both the topics using an unsupervised language modeling technique [16]. The vocabulary provided us a deeper understanding of the nature of the discussion that this community produced to support each other as a close-knit group. We found that posts and comments that are regarding symptoms were often asking some form of assurance from the community that their lingering, unusual symptoms were not completely unheard of by other members of the group., whereas the posts and comments that were not about symptoms often shared their journey to recovery to keep the hope up among the members of the community.

To our knowledge, this is the first study that has utilized social media discussions of the COVID long haulers' community to develop a visual framework for characterizing the symptoms reported by this community. In addition, our analysis has shown how social media groups might organically develop a support system when there is a scarcity of regular resources and infrastructure. Our findings have widespread implications for online community design and functioning, that can better support the needs of ad-hoc groups. This research also opens up new avenues for public health intervention and policy change to better address the requirements of minority groups during critical times such as a global pandemic.

Ethics and Disclosure. Because we used social media data of a private group, we had to get permission from the moderator of the group in order to get this study approved by the relevant institutional review board. In addition, we took great care in the way data and analyses are presented in the paper, for instance, by avoiding any personally identifiable information. We intentionally avoided including any quote in the paper as the data was collected from a "private" group and we wanted to ensure that the data is not traceable at all. No coauthors on the papers have experienced any long-hauler symptoms yet. We joined the group only as HCI researchers to collect data and later process them. We recognize and acknowledge the limitations of our methodological approach and our position as researchers and outsiders to this particular online community. We describe our limitations and ethical considerations further in the Discussion section.

2 RELATED WORK

2.1 Socio-economical Challenges during COVID

On March 11, 2020, World Health Organization (WHO) declared COVID-19 as a global pandemic. At that time, only 118,000 confirmed cases were reported which has increased to 187,519,798 so far and we are still counting. Because of this rapid increase in the number of confirmed cases, COVID-19 did not remain only as a regular infectious disease; rather it has impacted almost every aspect of our society. For instance, when counseling services to mental health patients were provided through an online program, counselors found it challenging to build rapport in the online environment. On the other hand, mental health patients faced hardship related to finance, housing, and distance learning due to the pandemic which often resulted in an increased level of anxiety, stress, addiction, depression, or psychosis [33, 48].

Getting accustomed to online programs and the working environment turned out to be challenging not only for mental health patients. Older adults also faced numerous challenges because of this change of norms. According to a 2017 Pew Research study, three-quarters of those older than 65 said they needed someone else to set up their electronic devices [1]. A third also said they were only a little or not at all confident in their ability to use electronics and to navigate the web [11]. This problem became much worse during the pandemic when older adults had to isolate themselves completely as they were facing a high risk of getting infected by the COVID-19 virus [34]. Similar to older adults, getting used to the online environment introduced many challenges for children as well. An initial survey has found that due to lockdown, children could rarely interact face-to-face with their teachers. In fact, some schools and families did not even have any dedicated resources and facilities for participating in online programs. The problem was more grave for below-average income families as children from those families felt less strongly about their own capacities to cope with online learning activities than other children [50].

Below-average income families also experienced a great deal of unemployment and lay off during this pandemic which made it even harder for their families to survive. In Joe Pinker's (2020) Atlantic essay entitled, "The Pandemic Will Cleave America in Two" [40], he highlighted two distinct experiences of the pandemic. Highly educated people would be able to continue working from home and would be able to avoid public gatherings as advised by CDC. However, this was not the case for low-income jobs. Unlike high-income jobs, the majority of the low-income jobs could not be done staying at home. These people either risked their life to continue working throughout the pandemic or lost their job during the pandemic because of downsizing. Unemployment was found to be at the core of many socio-economical challenges such as lower standard of living [5], domestic violence [28], and mental health issues [26].

All these socio-economical challenges discussed so far had hit even harder to the members of the minority community such as Black communities, Latinos, immigrants and so on [53]. In this paper, we focused on one such minority community — the COVID-19 long-hauler community. The long-hauler community was created as a direct impact of the pandemic. Until recently, even medical professionals knew very little about the conditions and challenges faced by the long-hauler community. Thus, the long-hauler community tried to comfort and assure each other by forming social media groups. In this paper, we focused our attention on one such group that was formed on Facebook to emotionally support COVID-19 long-hauler patients. We aimed to investigate from their discussions how they extended support to each other when only a few people even believed in their concerns.

2.2 Minority Groups on Social Media

In the last few years, many research studies have found a close connection between minority groups and social media. For instance, Muller et al. have shown that since the 2016 presidential primaries and President Donald Trump's political rise, one standard deviation increase of Twitter usage increases anti-Muslim hate crimes by 32% [35]. In fact, Trump's tweets about Islam-related topics predict increases in xenophobic tweets by his followers, cable news attention paid to Muslims, and hate crimes on the following days. During the global pandemic, another form of hate message that spread across all major social media platforms was Anti-Asian racial messages. About 17% of Asian Americans said in January 2021 they experienced severe online harassment compared with 11% during the same period last year. Online hate and harassment aren't unique to Asians. For many years, social media users who identify as Black, Jewish, transgender, or as part of other marginalized groups have also complained that Facebook and Twitter aren't doing enough to stamp out hate speech, despite having rules against that type of behavior [52].

Despite the threat of spreading hate messages, misinformation, disinformation, and rumors through social media, minority communities often found social media as the only place where they could express their opinion freely and fearlessly. A 2011 national survey of LGBTQ youth reported that this population spent more time online, and were more likely to have close online friends, compared to non-LGBTQ youth [18]. Social media platforms enable LGBTQ people to seek and find health information [21, 30], yet this practice can sometimes be invalidating when one's specific identity or health concern is not represented online [30]. Tumblr has often been recognized as particularly LGBTQ friendly [8, 9, 15, 38]. Some of Reddit's features, such as anonymous and pseudonymous identities, enable LGBTQ communities to form and thrivedarwin2017doing, farber2017transing. Overall, studies found that social media can be an important place for online LGBTQ presentation due to the ability to maintain boundaries between different identities and networks, thus enabling a relatively safe space for identity exploration and transition [6, 20].

In addition to identity exploration, minority communities often relied on social media to protest discrimination against the communities. In 2014, following the non-indictments of officers in the murders of Michael Brown and Eric Garner, the youth of color used hashtags such as "#AllLivesMatter" and "#BlackLivesMatter" to shape the national discourse about race in the wake of high-profile tragedies [7]. Black Lives Matter (BLM) group frequently used social media for building connections, mobilizing participants and tangible resources, coalition building, and amplifying alternative narratives [24, 36].

Similar to the BLM group and LGBTQ community, the long-hauler community also utilized Facebook to get united and to find strength and solidarity within the group. However, unlike other minority groups, we know very little so far about the long-hauler community. Our work aims to fill this existing gap by examing this community through their discussion on Facebook.

2.3 Social Discrimination against Minority Communities

Virtually all countries in the world have national or ethnic, linguistic, and religious minorities within their populations. Many violations of civil, political, economic, social, and cultural rights have a basis in discrimination, racism, and exclusion on the grounds of the ethnic, religious, national, or racial characteristics of the victim group. In the United States, discrimination against minority issues has been on the agenda for more than 60 years. Prior work has shown that individual and institutional measures, responsible for racial discrimination, were associated with the poor health status of Asian Americans [17]. In addition to the physiological health condition, in the USA, interpersonal discrimination has also been associated with increased rates of hypertension, depression, and stress; poorer self-rated health; and more reported days spent unwell in bed [3, 27].

Fig. 1. A word cloud showing all the major topics discussed in the COVID-19 Long-Haulers Discussion Group. We included all topics on which there were at least 10 posts posted on this group's page. The text size of the name of each topic is proportional to the number of posts on that topic. The list of topics includes several symptoms experienced by long hauler patients and many emotional states explaining the state of the mind of the members of this group.

Other than the health sector, discrimination against minority communities has also been observed in the job sector. Imana et al. [23] found that job ads delivered through Facebook were skewed for gender minorities and it could not be justified by differences in qualifications. Discrimination was also observed in the job sector when promotions and increases in salary were considered [12]. Personal prejudice is a critical issue that makes minority communities experience discrimination in many areas. For instance, careful observation revealed that rental housing discrimination still exists in several important types of housing agent behavior, and in most of the cases, the discrimination was caused by agents' own prejudice [10] and/or by the centrality of powerful institutional (i.e., banks, realtors, and insurance companies) [41].

In our paper, we focused on COVID-19 long-hauler community, a minority community that primarily experienced discrimination because no one knew how to explain such concerns. Together, this body of research (discussed in this section) motivated us to closely observe this community by following their private social media group. We believe our findings will add value to the existing literature on minority discrimination, especially to understand the characteristics of ad-hoc minority communities who may appear during any type of natural or man-made crisis.

3 METHOD

3.1 Collecting Social Media Data on COVID Long Haulers' Community

The first step in this phase was to select a social media group dedicated to long-hauler patients. There are several Facebook groups and Twitter communities that could satisfy this criterion. Some notable Facebook groups are COVID-19 Long-Haulers Discussion Group, COVID-19 Long Haulers Support, CoVid-19 Long Hauler Support, "Long-Haulers" Coronavirus Covid-19 Survivors Support Group, and so on. We decided to focus on the "COVID-19 Long-Haulers Discussion Group" for three reasons: 1) it was one of the earliest groups for long haulers' community on Facebook, 2) it was dedicated to long haulers primarily as the name suggests, and 3) it was a private group. We hypothesized that a private, fairly large social media group for long haulers' would make long

(a) (b)

Fig. 2. Example word-trees built around “covid” and “symptoms” on snippets of Facebook posts and comments in our dataset. The font sizes are proportionate with relative occurrence.

hauler patients feel more assured and comfortable to share their ideas, thoughts, experiences, and opinions. We acknowledge that **public** social media groups might have introduced different but equally important nuances to this study. We would leave it for future work. Figure 1 shows the list of all major topics that were mentioned in posts by the members of this group.

To analyze posts and comments available on this group, we manually copied all the posts made by the members of this group every day. The comments of some of the posts used to remain hidden when there were many comments under one post. We expanded all comments before copying them. We continued this process from Nov 3rd, 2020 till February 6th, 2021. In total, we collected 12,436 posts during this duration which contained 186,860 comments. On average, there were 15.02 comments per post. These posts were posted by 7,446 unique users (mean=1.67 posts per user). To give the reader a broad sense of our dataset, we present example word-trees in Figure 2 which show examples of post snippets of how people express about “symptoms” and “covid” on this group’s page on Facebook.

3.2 Adopting the Latent Dirichlet Allocation (LDA) technique to identify Main Discussion topics of the Group

We began by applying standard text-processing steps such as removing special characters, hyperlinks, punctuation, digits, stop words, and lowercasing all characters. We adopted the Latent Dirichlet Allocation (LDA) technique [4] to extract the range of discussion topics contributed by the Long haulers’ community through posts and comments. LDA is a widely used unsupervised statistical model to discover hidden topics by analyzing the semantic structure of the documents. Each topic consists of a set of keywords that define it, and text tokens are distributed over latent topics throughout each document. We treated each comment or post as a document and applied LDA on all of them.

The performance of the LDA model depends on the choice of hyperparameters α and β and the number of topics (k). Here, α controls the sparsity of document-topic distribution and β determines the sparsity of topic-word distribution. A low value of α is preferred (less than 1), because it produces a sparse distribution, leading to very few topic assignments per comment. This intuitively makes sense, because it is almost unlikely to mention a large number of topics in a single Facebook post or comment. Similarly, lower values of β favor fewer words per topic. To tune the value of the hyperparameters, we followed the similar procedure that was proposed by Pathik et al. [4]. We considered k = 20 topics as a seed value and ran the LDA model for a range of values of α and β. We considered all values in the range of [0.01 0.99] at regular intervals of 0.05. For each unique combination of α and β, we ran the model and recorded the coherence score. Thus, we chose α =

0.01 and $\beta = 0.11$ as the best-fitting hyperparameters for our dataset since the coherence score of the model was the highest for this combination.

Once the values of α and β were identified, we followed the same procedure to tune the value of the number of topics (k). With $\alpha = 0.01$ and $\beta = 0.11$, we ran the model for all values in the range of [5 50] at regular intervals of 5. We observed the highest coherence score at k = 15 and the score did not increase significantly after that. We also investigated the topics themselves and increasing the value of k beyond 15 resulted in repeated appearances of the same keywords in multiple topics which were not intended in our context. Finally, we decided to run the model for $\alpha = 0.01$, $\beta = 0.1$, and k = 15 and generate topics for further analysis. Once the topics were identified, two human coders familiar with the concept of social media group for COVID long haulers independently reviewed those 10 topics and the top words in each topic. Following an inductive open coding method, they individually identified the non-overlapping themes from those topics. In the process, they merged two or more topics when they were thematically overlapping with each other. Finally, they resolved disagreements through discussions. We identified two main themes.

Table 1 lists the ten topics presenting the main discussions in the long-haulers' Facebook group. The table shows the percentage of posts (column 2) on each topic along with the top five keywords. Primarily, the topics generated by LDA can be divided into two major categories: 1) the topics where group members discussed a specific symptom or a group of symptoms that they were experiencing regularly and 2) topics that were not directly related to any symptoms. Members of the group discussed various ongoing symptoms that they experienced for months after COVID. Some notable ones were a large amount of hair loss, chest pain, difficulty in breathing, abnormally high or low blood pressure level, brain fog, fatigue, and loss of smell and taste. They also discussed how some of these symptoms continued for months and on many occasions, even their doctors failed to treat those symptoms.

The topics that were not related to symptoms mostly highlighted community-wise collaboration and empathy. Members of the community expressed condolences when someone went through a tough time. They also expressed their warm gratitude to the group for support during the hour of need. The timeline of these posts and comments were just before the time when the COVID vaccine initially became available for everyone. The members of long-haulers' groups consulted each other to get additional information about the vaccines. This indicates a sense of trust among the members who were essentially ignored for a long-time by the healthcare system.

Overall, the main discussion topics identified by LDA show that the members of the group considered this space as a trusted place where they felt comfortable to discuss their lingering symptoms, vulnerabilities, and challenges. They also found this group useful to gain assurance that they were not experiencing something alone; rather as a group, they were facing these conditions which might have given them a sense of hope and courage to face these conditions as a strongly bonded group.

To further investigate the minute nuances of these discussion posts and comments, we conducted qualitative annotation. Next, we explain that process.

3.3 Annotating Long-Hauler discussion on Facebook

To make sure that we did not miss any critical discussion topics (other than those that were identified by the LDA technique), we performed manual annotation of a subset of our dataset. Two human coders (one of them was the first author of this paper) who were familiar with the COVID long hauler community and their activities on social media examined a random sample of 1000 entries from our dataset. In the absence of labeled ground-truth data, they adhered to an open inductive coding approach [18]. During this coding process, we organized three brainstorming sessions where both

Table 1. LDA table with 10 topics. Column 2 shows the percentage of comments in that specific category. Column 4 lists five representative words from each topic.

	Lexical Group	Percentage of Posts/Comments	Top Five Keywords
Topics related to symptoms	Symptom of Hair Fall	5.8%	“hair”, “falling”, “losing”, “plasma”, “convalescent”
	Symptoms Reappearing for Months even after negative test result	17.87%	"symptoms", "months", "still", "tested, "negative"
	Symptoms of Loosing Smell and Taste	7.3%	“smell”, “feel”, “taste”, “back”, “sinus”
	Symptom of Irregular Heart Rate and Blood Pressure	15.8%	“heart”, “blood”, “rate”, “pressure”, “low”
	Symptoms of Itchiness and hives	12.1%	“itchy”, “hives”, “phantom”, “bumps”, “sore”
	Symptoms of Pain	9.2%	“pain”, “chest”, “neck”, “else”, “cough”
Topics not related to symptoms	Sharing News about Recovery and Returning home	7.1%	“hospital”, “home”, “going”, “work”, “relief”
	Seeking suggestion for vaccination	10.9%	“vaccine”, “dose”, “anyone”, “recommendation”, “received”
	Pray for loved ones recovery and offering condolences	6.03%	“family”, “prayers", “god”, “please”, “group”
	Thanking the community	9%	“thank”, “grateful”, “prayer”, “wishes”, “luck”

coders discussed their preliminary thoughts with each other. We followed an iterative process and after multiple iterations, we identified 24 initial themes.

Next, to avoid any bias imposed by the first author of this paper and to make this annotation process more applicable in general, we invited five undergraduate students, all with backgrounds in social media data analysis, social science, and behavioral psychology, to examine another random sample of 500 comments. This new set did not contain any comments from the previous set. To provide background on the annotation process, we conducted an hour-long information session that involved discussing themes identified earlier along with specific example comments. Following this discussion, all coders independently coded the new set of 500 comments. They could either apply any theme from the existing pool of 24 themes (if applicable) or create a new theme for each comment based on their judgment. Finally, we discussed their coding experiences and received feedback about potentially ambiguous, misrepresented themes, and possible new themes.

Based on the discussion with undergrad coders, we modified, removed, and added a few themes. Next, to assess the effect of the changes, the first author and a social science expert coded another random sample of 600 comments (did not include comments from any previous set). The disagreements in annotations were resolved through discussion until consensus was reached. We also combined multiple initially identified themes that were closely related to each other. Finally, we achieved a

substantial agreement based on Cohen's kappa test (K = 0.84). Combined efforts in the three stages resulted in the same two major themes (as we identified from LDA analysis): 1) topics related to symptoms and 2) topics not related to symptoms. However, in the "topics not related to symptoms" theme, we identified the following four new sub-categories: 1) expressing frustration, embarrassment for nightmares, memory loss, experiencing symptoms on some days, 2) experience panic attacks because of an anticipation of some irreversible damage, 3) overwhelmed with financial instability, and 4) seeking suggestion.

3.3.1 Expressing Frustration and Embarrassment for Nightmares, Memory Loss, Experiencing Symptoms on Some Random Days. Other than reporting physiological symptoms, many members of this group discussed having vividly realistic nightmares regularly (N=32). On some occasions, members of this group claimed that they had got over all other symptoms but had harrowing nightmares almost every night which made them even scared to go to sleep. Those nightmares were frequently about getting infected with the COVID-19 virus again. Many members with such conditions felt embarrassed to share this condition with their primary care physician assuming that they would make fun of them or would not take their condition seriously. Another major reason for such an emotional outbreak was due to brain fog or temporary memory loss which often made their daily tasks increasingly difficult to accomplish (N = 41). As someone described that in the middle of a phone call, he forgot what he was talking about, and out of embarrassment, he had to hang up the phone on his best friend. In another post, someone discussed how she forgot simple things about their household which made it extremely challenging to complete simple household chores. Finally, a large number of members of this group got frustrated since their symptoms were coming back on random days, and on other days, they were perfectly normal. Not being able to anticipate how the next day or next week would look like made them frustrated and helpless. Overall, comments and posts in this category could not be marked as COVID long-hauler symptoms (at least those who described these experiences felt like that), but these experiences made them feel exhausted and drained on regular basis.

3.3.2 Experience Panic Attacks because of the Anticipation of Irreversible Damages. Some members of this group did not talk about any specific lingering symptom but shared their thoughts of experiencing panic attacks frequently. Some members believed that COVID-19 has changed their genetic structures irreversibly. For instance, one member of the group stated that even after fully recovering from COVID 10 months back, she could never be as active and as energetic as she was before. She believed that COVID had affected her genes and although she was not having any symptoms and tested negative, she believed that her quality of life had degraded significantly because of the infection. Other members (N=9) also suspected of the change of genetic structure and claimed that they had constant panic attacks thinking that because of this change, they would be more likely to experience chronic illnesses such as diabetes, high blood pressure, and autoimmune diseases. Few members of the community (N=14) shared their experiences of seeing close family members going through extreme conditions because of COVID. They claimed that they had panic attacks thinking that their family members were critically ill because of certain properties of their genes and if somehow they get infected by the COVID virus, their condition would also be critical because they all had similar genetic structures.

3.3.3 Overwhelmed with the Financial Instability. Concerns about financial instability were another topic that was discussed in this group frequently. Many members of the group (N = 38) reported being laid off from their jobs because of their lingering symptoms. Some of them (N=9) were tensed because they felt that the reason for their getting laid off would always be there in their professional record and that would potentially be considered as a negative point when would apply for new jobs

Table 2. Median metrics in k-fold (k=5) cross-validation.

Model	Accuracy	Precision	Recall	F1	AUC
Naïve Bayes	0.788	0.799	0.947	0.867	0.657
Linear Regression	0.764	0.790	0.919	0.850	0.636
SVM	0.767	0.741	0.958	0.868	0.510
Random Forest	0.827	0.817	0.978	0.894	0.677
XGBoost	0.853	0.880	0.931	0.905	0.775
CNN	0.614	0.686	0.722	0.749	0.535

in the future. Some members (N=15) expressed their fear that they would not be able to perform like they used to do before COVID because they felt that they could not recover as much as they expected. Long hauler patients were also deeply concerned that their employer would not believe in their lingering symptoms because they did not have any medical record to prove their symptoms were legit. Some members even mentioned that they hid their sufferings from everybody in their workplace as much as possible because they were afraid that their employer and co-workers might think that their conditions were all psychosomatic symptoms and that thought might impact their future promotions and salary raises in the company.

3.3.4 Seeking Suggestions. (both own work and for partner's work who died) Members of this social media group trusted each other and thus, they asked for suggestions from the members of the group on various topics. LDA technique identified posts and comments where the members of the group sought suggestions on vaccination, its effectiveness, and risk factors. Although vaccination was one of the most popular topics for seeking suggestions, that was not the only topic on which people asked for suggestions from each other. Another popular topic on which members of this group frequently asked suggestions was outreach programs for COVID-19 patients or especially for COVID-19 long-hauler patients. Other popular topics in this category were cures for unusual symptoms, various information related to health care facilities, accessibility of certain products which were hard to find during the initial stage of the pandemic.

In summary, qualitative annotation allowed us to construct a more clear picture of this long-haulers' social media community. Manual annotation highlighted critical nuances of their discussions among the members of the group which were not captured by the LDA technique. However, this analysis did not identify any conceptually new theme from our dataset. Thus, we applied machine learning algorithms to automatically classify all entries (both posts and comments) in our dataset as to one of these two classes.

3.4 Building A Machine Learning Classifier for Identifying Symptom Related Posts and Comments

Our next goal centers around identifying social media posts and comments discussing COVID-19 long-hauler symptoms at scale. We draw on natural language analysis techniques to build a machine learning classifier on the annotated dataset. We describe our approach, features, and models below.

3.4.1 Machine Learning Features. Inspired from several prior works in social media language [25, 43, 44], our work uses four kinds of features:

Latent Semantics (Word Embeddings). To capture the semantics of language beyond raw keywords, we use word embeddings, which are essentially vector representations of words in latent semantic dimensions. Several studies have revealed the potential of word embeddings in improving natural language analysis and classification problems [32]. In particular, we use pre-trained word embeddings

Table 3. Incremental accuracy metrics of adding features in the COVID-19 Long-Hauler Classifier (AdaBoost).

Model	Accuracy	Precision	Recall	F1	AUC
N-grams	0.724	0.770	0.853	0.813	0.632
.+ Word Embeddings	0.778	0.785	0.895	0.855	0.660
.+.+Sentiments	0.793	0.802	0.939	0.865	0.690
.+.+.+LIWC	0.853	0.880	0.931	0.905	0.775

(GloVe) in 50-dimensions that are trained on word-word co-occurrences in a Wikipedia corpus of 6B tokens [39].

Psycholinguistic Attributes (LIWC). Prior literature in the space of social media and psychological wellbeing has established the potential of using psycholinguistic attributes in building predictive models [14, 43] We use the Linguistic Inquiry and Word Count (LIWC) lexicon to extract a variety of psycholinguistic categories (51 in total). These categories consist of words related to affective process, cognitive and perceptual process, informal language, time orientations, linguistic dimensions, biological process, social process, and personal concerns [49].

Open Vocabulary (n-grams). Drawing on prior work where open-vocabulary based approaches have been extensively used to infer psychological attributes of individuals [45, 46] we also extracted the top 500 n-grams (n = 1, 2, 3) from our dataset as features.

Sentiment. An important dimension in social media language is the tone or sentiment of a post. Sentiment has been used in prior work to understand psychological constructs and shifts in the mood of individuals [19, 42]. We use Stanford CoreNLP's deep-learning based sentiment analysis tool [31] to identify the sentiment of a post among positive, negative, and neutral sentiment labels.

Modeling Approach. We used the 600 manually annotated posts and comments from the previous section to build a machine learning classifier with a total of 619 features, as described above. We consider and evaluate multiple classifiers, including Naive Bayes, Logistic Regression, Random Forest, Support Vector Machine (SVM), AdaBoost, and Convolutional Neural Network (CNN) algorithms. We use stratified k-fold cross-validation (k = 5) to parameter tune our classifiers. Table 2 summarizes the performance metrics of these models. All of these classifiers performed better than the baseline accuracy of 50% on our dataset (based on a chance model). We found that the Boosting classifier outperforms all with a median AUC of 0.78, median precision of 0.88, and median recall of 0.93. Table 4 summarizes the performance metrics of this classifier, where we find that the classifier is reasonably stable (STDEV = 0.03) across the five folds, and Table 3 summarizes the step-wise improvement with the addition of each kind of feature in the AdaBoost model. For the rest of the paper, we would use the AdaBoost as the classifier for finding posts and comments on COVID-19 long-hauler symptoms.

We used the classifier for long-hauler symptoms to label all our 199,296 posts and comments of our dataset. We found that 68% of them (135,521) were predicted to be about some kind of long-hauler symptoms. This section would first analyze the linguistic markers associated with long-hauler symptoms. The linguistic markers would show more fine-grained nuances based on the regular conversation of the social media group of the long-haulers' community. Then, we illustrated all major symptoms reported in our dataset using a network-graph (symptoms that were reported at least 10 times in our dataset). Doing so, we would essentially present a vivid picture of the range of

Table 4. Detailed accuracy metrics in k-fold (k=5) cross-validation in the COVID-19 Long-Hauler Classifier (AdaBoost).

Metric	Min.	Max.	Mean	Stdev.
Accuracy	0.750	0.853	0.783	0.029
Precision	0.788	0.880	0.843	0.016
Recall	0.876	0.931	0.913	0.033
F1	0.830	0.905	0.865	0.020
AUC	0.666	0.775	0.696	0.030

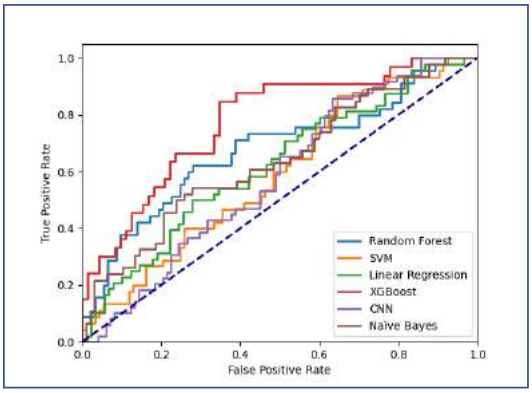


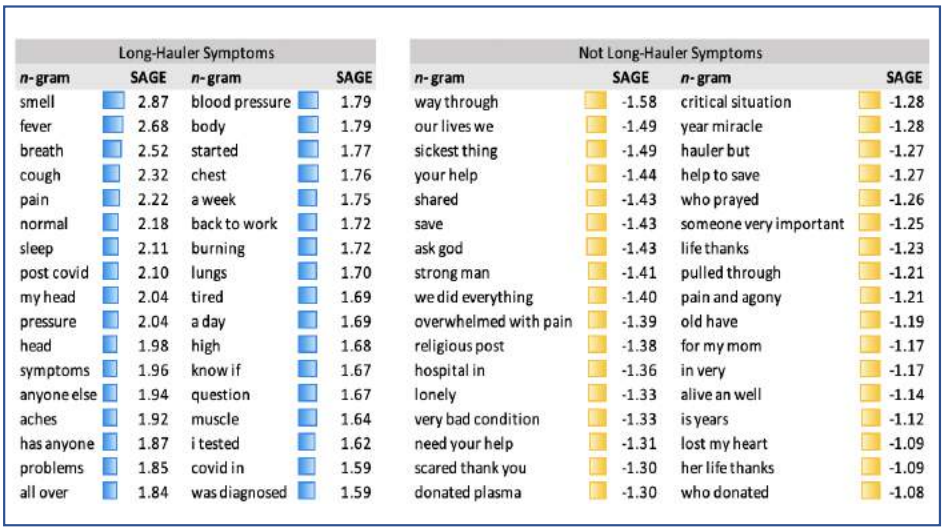
Fig. 3. COVID-19 Long-Hauler Classifier accuracy metrics on k-fold cross validation: AUC curve

symptoms that long-hauler members are facing continuously where some of these conditions were completely ignored by health care providers until recently.

3.5 Analyzing Language Cues Associated with Social Media Discourse of Long-Haulers' Community

3.5.1 Finding discriminating language cues. We first examined the language markers associated with the symptoms reported by the long-haulers' community. We employed an unsupervised language modeling technique known as the Sparse Additive Generative Model (SAGE) [16]. Given any two documents, SAGE selects discriminating keywords by comparing the parameters of two logistically parameterized multinomial models, using a self-tuned regularization parameter to control the trade-off between frequent and rare terms. We used SAGE to identify discriminating n-grams (n=1,2,3) between the posts and comments that discussed long-hauler symptoms and those that did not discuss long-hauler symptoms. The magnitude of the SAGE value of a linguistic token signals the degree of its "uniqueness", and in our case a positive SAGE (more than 0) indicates that the n-gram is more representative for long-hauler symptoms, whereas a negative SAGE denotes greater representativeness for its absence.

3.5.2 What do the discriminating keywords say? Figure 4 reports the top 34 discriminating keywords that occurred in posts and comments related to symptoms and not related to symptoms. We found that keywords to the symptom category contained almost all the keywords that were somehow related to any symptom. Other than symptoms, some notable keywords in this category are "has anyone" and "anyone else". A majority of the members initially joined this group to ensure that they



Long-Hauler Symptoms			
n-gram	SAGE	*n*-gram	SAGE
smell	2.87	blood pressure	1.79
fever	2.68	body	1.79
breath	2.52	started	1.77
cough	2.32	chest	1.76
pain	2.22	a week	1.75
normal	2.18	back to work	1.72
sleep	2.11	burning	1.72
post covid	2.10	lungs	1.70
my head	2.04	tired	1.69
pressure	2.04	a day	1.69
head	1.98	high	1.68
symptoms	1.96	know if	1.67
anyone else	1.94	question	1.67
aches	1.92	muscle	1.64
has anyone	1.87	i tested	1.62
problems	1.85	covid in	1.59
all over	1.84	was diagnosed	1.59

Not Long-Hauler Symptoms			
n-gram	SAGE	*n*-gram	SAGE
way through	-1.58	critical situation	-1.28
our lives we	-1.49	year miracle	-1.28
sickest thing	-1.49	hauler but	-1.27
your help	-1.44	help to save	-1.27
shared	-1.43	who prayed	-1.26
save	-1.43	someone very important	-1.25
ask god	-1.43	life thanks	-1.23
strong man	-1.41	pulled through	-1.21
we did everything	-1.40	pain and agony	-1.21
overwhelmed with pain	-1.39	old have	-1.19
religious post	-1.38	for my mom	-1.17
hospital in	-1.36	in very	-1.17
lonely	-1.33	alive an well	-1.14
very bad condition	-1.33	is years	-1.12
need your help	-1.31	lost my heart	-1.09
scared thank you	-1.30	her life thanks	-1.09
donated plasma	-1.30	who donated	-1.08

Fig. 4. Top discriminating n-grams (n=1, 2, 3) in posts and comments with and without COVID-19 long-hauler symptoms (SAGE Analysis [16])

were not the only people who were suffering from unexplained, lingering symptoms for days or even for months. Thus, these are the common phrases used frequently in posts. Another important keyword in this category is “back to work”. Many members of the long-haulers community were highly concerned about going back to their work as they were not feeling completely recovered but they could not produce any medical document saying that their condition and symptoms were real.

In contrast, the keywords in the other category did not highlight any symptoms; rather it contained phrases on a range of emotional expressions. For instance, we noticed that keywords such as “who prayed”, “who donated”, “life thanks” were used to show thankfulness to the community members who supported and prayed for each other at the time of crisis. Other groups of keywords that are frequent in this category are “sickest thing”, “we did everything”, “very bad condition”, and “critical situation”. These keywords were used mostly to share sufferings and challenges that long-hauler patients faced and to express their feelings of utter helplessness that they had to deal with throughout. These posts could be associated with the fact that this social media group helped their members to feel like a part of the community and not being left out anymore which was crucial at the time when no one believed in them.

3.6 Visualizing Long-Hauler Symptoms using Network Graph

Our step-by-step analysis of long-haulers’ social media community has shown that querying about different types of lingering but unusual symptoms was the single most important topic in this group. Our best performing machine learning classier (AdaBoost) identified that more than two-thirds (68%) of the posts and comments belonged to the symptom category. Identification of these large number of symptoms leads us to ask the following question: how were these symptoms related to each other and which were the most frequently appeared symptoms?

To visualize the interconnection among symptoms, we decided to create a network-graph only for long-hauler symptoms. To this end, we extracted all the keywords from the dataset and information about their co-occurrences. But the raw dataset consists of many meaningless words, punctuation, offensive words, and symbols. In our case, those acted as noise and did not contribute any information

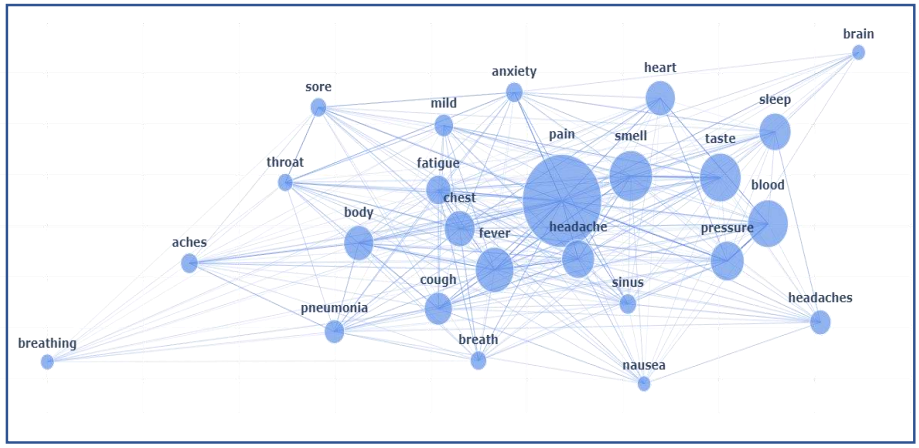


Fig. 5. A network graph showing COVID-19 Long-Hauler symptoms extracted from social media (Facebook) group's posts and comments. The graph shows the top 25 symptoms for better visibility. In total, we identified 264 symptoms. The full network graph is attached in the Appendix A.

in building the network graph. So, we used a pattern/rule-based strategy to extract causal links applying an NLP framework that could use syntactic information. The first step was to pre-process the raw data and extract only the useful information. Once the noise had been removed, the next step was Part of Speech (PoS) tagging. A PoS tagger's purpose is to assign a linguistic tag or information to tokens. We filtered in only such tagged terms because we were only interested in retrieving the symptom element of the text, which was typically forms of nouns or adverbs (primarily).

Our process became more efficient step by step. We used a dictionary of describing the symptom-related keywords, words were chosen from that, which were all nouns. The primary focus was on unigrams. As a result, removing all Noun-Noun (NN) words would direct us to the synonyms that were needed. The PoS tagger NN was implemented. It was not, however, 100% accurate. To increase the overall performance, we included bigrams and further n-grams because the symptoms can be a group of words. Bigrams or further n-grams could also be used to express adjectives for a specific symptom, such as chest pain or awful taste. This highlighted the importance of a model that could handle n-grams larger than unigrams.

In addition, noise removal was ineffective in retaining only the noun parts of the input data. By using n-grams we built a pattern to identify the symptom-related keywords which were termed as cause-effect relation extraction. Cause and effect relationships could be expressed in a variety of ways, including verb phrases and noun phrases. The cause-effect relations were determined based on trigger terms. For example, consider a sentence: "I have lost taste and smell", we can see the words "taste", "smell" are nouns that are combined by conjunction constitutes the symptom dictionary. We needed to build a pattern-based method that extracts based on the part of speech of a particular word in the sentence. By using the PoS tagger and the trigram we should successfully extract these keywords.

Figure 5 shows the network graph drawn with the most frequent 25 symptoms from our dataset. The full list contains 264 symptoms in total. Since the full graph is hard to read, we decided to presents the top 25 symptoms in this graph. For your reference, the full graph is included in Appendix A. As the graph shows, the most frequently appearing symptom is "pain" which co-appeared with several other major symptoms such as fever, headache, smell, heart, and sinus. This network graph shows that many symptoms co-appeared in the social media data of the community. Since many of these symptoms are common Flu-like symptoms, that might be one possible reason why health care providers initially did not consider them seriously and in many scenarios, just did not take any action at all.

4 DISCUSSION

4.1 Theoretical Implications

This study assesses long haulers' community and their harrowing journey through difficulties via an inexpensive and unobtrusive data source, social media data. Long hauler patients received denials and ignorance throughout the pandemic. Only recently, initiatives are being taken to establish facilities that would be equipped to assist conditions related to long-hauler patients. However, gathering information from existing long hauler patients and taking action as per the requirements would most likely take a substantial amount of time and effort. Since social media data consists of long hauler patients' self-initiated and candid opinions and experiences, this data provides us with a rich and accessible lens to examine this community, their requirements, and complaints. This approach has clear advantages beyond traditional mechanisms — such as surveying a large group of COVID-19 patients and closely monitoring their symptoms and lingering conditions. The purpose of this study is to not replace the traditional approach; rather establish a complementary strategy so that new infrastructures getting built to assist long haulers can make the best use of social media data. The large-scale availability of social media data provides opportunities to understand the breadth of mental, physiological, socio-economical issues that the long-haulers' community is facing throughout this global pandemic.

Summarily, this study opens up interest in conceptualizing long hauler patients specifically (their unique conditions) without confusing them with regular COVID patients or people who faced extreme socio-economical hurdles due to the pandemic. In particular, the in-depth linguistic analysis allowed us to observe the discussion that revealed not only a long list of lingering symptoms but also mental and social issues which are often missed. The network graph visualization illustrates the importance of studying these conditions as a group of co-related symptoms rather than any separate ailment.

During the pandemic, social media was frequently accused of spreading fake news, misinformation, and disinformation on mask usage, the safety of getting vaccinated, COVID-19 prevention measures and treatments. In recent work, Su et al. [54] found that social media news use was associated with higher conspiracy beliefs. Individuals who trust social media news more are more likely to believe in conspiracy theories. Our findings present an opposite side of social media where a minority community (COVID long haulers) gathered together and found assurance, hope, and support from each other. This private Facebook group provided them a trusted place where no one was dismissive of their conditions; rather the community made them feel visible, understood, and believed.

4.2 Practical Implications

4.2.1 Monitoring the existing symptoms and their continuous changes over time. The challenges and discrimination against minority groups such as COVID long haulers' community are often difficult to assess at a finer granularity. Although long hauler patients did not experience denials and negligence for a long time (as other minority communities such as the LGBTQ+ community), many

of them had to go through physiological and mental challenges that they had no prior experience with. They had no way to prepare for these challenges beforehand. Existing health care facilities often did not have the resources to assist long-hauler patients. Realizing these challenges, a Facebook group called “Survivor CORPS" has recently created a live guideline on how to establish and operate a multi-disciplinary Post-COVID Care Center [13] based on published practices of established centers from around the world. Our findings can provide valuable insights for such organizations, allowing a richer and nuanced understanding of long haulers’ physiological and mental conditions. The techniques followed in our analysis can help organizations build tools that can track long-hauler communities on a continuous and real-time basis to establish their initial base and later to evolve with the changing need and situation. A tool that can provide a continuous update based on social media data can be an excellent resource for keeping track of conditions and symptoms of a diverse group of patients who are hard to access in any physical location.

4.2.2 Developing Infrastructure for Long-Hauler Rehabilitation. Moreover, the symptoms experienced by long hauler patients often continue for a few weeks to several months. Our results have shown that lingering symptoms not only impact the physiological and mental conditions of the patients but also affect their overall lifestyle. Individuals going through such conditions often become temporarily incapable of continuing their existing jobs. Members from the long-haulers’ community also shared incidents where they had to quit their job because of some chronic condition such as ME/CFS (Myalgic encephalomyelitis/chronic fatigue syndrome) which can develop as an aftereffect of COVID-19 infection [22]. Since COVID long-hauler research is still at a very early stage, it is hard to predict how fast long-hauler patients would recover from this condition. In fact, in the last few years, the logic behind traditional physiological and cognitive therapy for ME/CFS has also received massive criticism. We believe our findings from posts not related to symptoms will provide initial insights for establishing rehabilitation facilities appropriate for long-hauler patients. Future work could adapt our approach to looking at more nuanced conceptualizations of long-term socio-economical challenges of long hauler patients that can be used to develop a concrete structure of rehabilitation program for them.

4.3 Policy and Social Implication

Pandemic generally occurs infrequently. The last major pandemic in the United States was H1N1 flu or commonly known as swine flu. It was detected in the spring of 2009 and 12,469 individuals died from the infection. In comparison to swine flu, we have seen 622,825 deaths so far in the USA because of COVID-19 which is approximately 50 times higher than swine flu. Preparing for such a major pandemic is always a challenging task but not quite impossible. Some notable examples are Singapore, Japan, Hong Kong, Taiwan, and South Korea. These countries severely suffered from SARS-COV-2 in 2003 and swine flu in 2009. However, they applied lessons learned from those outbreaks to revise their public health systems to such an extent that they could handle COVID-19 much better than the United States, some European countries such as Italy, and India [51].

Some of these statistics mentioned above might sound demoralizing but we can still learn from this pandemic. World Health Organization (WHO) has shown that infectious diseases outbreaks are emerging alarmingly regularly over the last 30 years [47] and we cannot deal with them just by luck. In the USA, we need to reorganize our public health care system significantly to be better prepared for the next possible pandemic. Our findings reveal that medical gaslighting might not happen only to women, black people, and Latinos. Without careful monitoring and a well-structured public health care system, this kind of discrimination can happen to anybody. A pandemic similar to COVID-19 can make such a condition significantly worse especially when the health care system is not prepared in advance. We believe our work will provide data-driven insights for informed policy decisions

and aid in building backup facilities even for people whose conditions are not initially well-defined by medical science. Layered, people-oriented health care divisions might become a more effective solution in these scenarios. Our findings have shown that automatic tools based on social media data can be developed to monitor these issues closely. Early detection of such conditions might be easier through regular monitoring of social media data where people from diverse communities are organically gathering together to share their experiences. To this end, we recognize that some of the alarms raised on social media can turn out to be false alarms and prompt actions against those alarms can be wasteful. We, therefore, suggest carefully constructed layered infrastructure which can filter false alarms raised on social media (if any) and pursue further actions only on legitimate cases.

From a methodological perspective, we further recognize the ethical complexities associated with automatically monitoring people's social media data for taking critical decisions in the public health care sector. The concern becomes much stronger when we are discussing private social media groups whose content should not be accessible to a non-member of the group. The factors that motivate users to express their opinion on social media and enable their candid self-disclosure may be confounded with their perceptions of being monitored. Moreover, people may have reservations as to who uses the results of such analysis — they may not be comfortable having government officials assess their social media data, as it can raise questions surrounding the privacy of social media data. If such a monitoring system is set without careful consideration of the privacy of the social media data, users of such groups might get discouraged and uncomfortable sharing their personal, sensitive information in the group.

The above factors and their potential risks and benefits need to be carefully evaluated before establishing infrastructure for continuous monitoring of online social media data.

5 LIMITATIONS AND FUTURE DIRECTIONS

We acknowledge that our work has limitations, many of which suggest interesting directions for future research. We do not make any population-centric assessments because the Facebook group considered in our work cannot be considered wholesome of online discussions of the long-hauler community. Rather, our work should be seen as a proof-of-concept study to examine the long-hauler community on social media. Future work that makes population-centric assessments associated with long-hauler patients should consider the caveats concerning missingness and quality of social media datasets.

Our work inherently suffers from self-selection biases, that it only works on the language of the individuals who self-selected to express themselves on online communities, particularly those that did not feel shy to share their experiences on social media. Relatedly, we only study the language on social media. Incorporating other behavioral and communicative signals like frequency of posting, the topic of interest, and the support-seeking or support-giving nature of posting can help us to comprehensively understand the long-hauler community on social media. Future work can further investigate this community across other mediums and social media platforms.

The machine learning classifiers can be further improved with more sophisticated models of machine learning and natural language processing. This can be tuned with respect to the objective of the problem, where our objective was to balance between predictability and interpretability — i.e., to not only build a stable model that reveals the potential in machine learning to scalably infer the language of Covid-19 long-haulers, but also to help us understand the linguistic nuances in expressing minority community on social media.

6 CONCLUSION

This paper studied the long-hauler community from the discussion of their private social media group. Adopting a combined qualitative and computational approach, this paper examined the

language on the online community specifically created for COVID-19 long-haulers, and makes three primary contributions. First, we identified two primary themes that can broadly summarize all the discussion topics of the COVID-19 long-haulers' Facebook group. Second, a machine learning classifier to identify social media posts discussing long-hauler symptoms at scale. The classifier used a variety of features, ranging across word embeddings, psycholinguistic attributes, sentiment, and open-vocabulary based n-grams. We achieved a mean AUC of 0.78. Finally, we conducted a deeper analysis on long-hauler community discussion to obtain lexicons of linguistic markers. We believe our work bears the potential to better understand the long-hauler community from an honest point of view especially when the community was ignored initially by health care providers. Our work also supports tailored public health intervention and policy change to better address the requirements of minority groups during critical times such as a global pandemic.

REFERENCES

[1] Monica Anderson and Andrew Perrin. 2017. Barriers to adoption and attitudes towards technology. https://www.pewresearch.org/internet/2017/05/17/barriers-to-adoption-and-attitudes-towards-technology/.

[2] Beth P Beckman. 2020. COVID-19: never seen anything like this ever! *The Journal of nursing administration* 50, 6 (2020), E3.

[3] Lisa F Berkman, Ichirō Kawachi, and M Maria Glymour. 2014. *Social epidemiology*. Oxford University Press.

[4] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022.

[5] David L Blustein, Ryan Duffy, Joaquim A Ferreira, Valerie Cohen-Scali, Rachel Gali Cinamon, and Blake A Allan. 2020. Unemployment in the time of COVID-19: A research agenda.

[6] Yuliya Cannon, Stacy Speedlin, Joe Avera, Derek Robertson, Mercedes Ingram, and Ashely Prado. 2017. Transition, connection, disconnection, and social media: Examining the digital lived experiences of transgender individuals. *Journal of LGBT Issues in Counseling* 11, 2 (2017), 68–87.

[7] Nikita Carney. 2016. All lives matter, but so does race: Black lives matter and the evolving role of social media. *Humanity & Society* 40, 2 (2016), 180–199.

[8] Andre Cavalcante. 2019. Tumbling into queer utopias and vortexes: Experiences of LGBTQ social media users on Tumblr. *Journal of Homosexuality* 66, 12 (2019), 1715–1735.

[9] Alexander Cho. 2018. Default publicness: Queer youth of color, social media, and being outed by the machine. *New Media & Society* 20, 9 (2018), 3183–3200.

[10] Seok Joon Choi, Jan Ondrich, and John Yinger. 2005. Do rental agents discriminate against minority customers? Evidence from the 2000 Housing Discrimination Study. *Journal of Housing Economics* 14, 1 (2005), 1–26.

[11] Kate Conger and Erin Griffith. 2020. As Life Moves Online, an Older Generation Faces a Digital Divide. https://www.nytimes.com/2020/03/27/technology/virus-older-generation-digital-divide.html?action=click&module=RelatedLinks&pgtype=Article.

[12] Courtney Connley. 2021. Why Black workers still face a promotion and wage gap that's costing the economy trillions. https://www.cnbc.com/2021/04/16/black-workers-face-promotion-and-wage-gaps-that-cost-the-economy-trillions.html.

[13] Survivor CORPS. 2021. Post-COVID Care: Guidelines for Multidisciplinary Care Centers. https://static1.squarespace.com/static/5e8b5f63562c031c16e36a93/t/605a8a3262f0191b99584df0/1616546355297/PCCC+Standard+of+Practice+3_23.pdf.

[14] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting depression via social media. In *Seventh international AAAI conference on weblogs and social media*.

[15] Stefanie Duguay. 2016. "He has a way gayer Facebook than I do": Investigating sexual identity disclosure and context collapse on a social networking site. *New media & society* 18, 6 (2016), 891–907.

[16] Jacob Eisenstein, Amr Ahmed, and Eric P Xing. 2011. Sparse additive generative models of text. In *Proceedings of the 28th international conference on machine learning (ICML-11)*. Citeseer, 1041–1048.

[17] Gilbert C Gee. 2008. A multilevel analysis of the relationship between institutional and individual racial discrimination and health status. *American journal of public health* 98, Supplement_1 (2008), S48–S56.

[18] Barney G Glaser, Anselm L Strauss, and Elizabeth Strutzel. 1968. The discovery of grounded theory; strategies for qualitative research. *Nursing research* 17, 4 (1968), 364.

[19] Scott A Golder and Michael W Macy. 2011. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science* 333, 6051 (2011), 1878–1881.

[20] Oliver Haimson. 2018. Social media as social transition machinery. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–21.
[21] Blake Hawkins and Jen Jack Gieseking. 2017. Seeking ways to our transgender bodies, by ourselves: Rationalizing transgender-specific health information behaviors. *Proceedings of the Association for Information Science and Technology* 54, 1 (2017), 702–704.
[22] HealthAffairs. 2021. Paradigm Lost: Lessons For Long COVID-19 From A Changing Approach To Chronic Fatigue Syndrome. https://www.healthaffairs.org/do/10.1377/hblog20210514.425704/full/.
[23] Basileal Imana, Aleksandra Korolova, and John Heidemann. 2021. Auditing for Discrimination in Algorithms Delivering Job Ads. In *Proceedings of the Web Conference 2021*. 3767–3778.
[24] Jelani Ince, Fabio Rojas, and Clayton A Davis. 2017. The social media response to Black Lives Matter: How Twitter users interact with Black Lives Matter through hashtag use. *Ethnic and racial studies* 40, 11 (2017), 1814–1830.
[25] Kokil Jaidka, Salvatore Giorgi, H Andrew Schwartz, Margaret L Kern, Lyle H Ungar, and Johannes C Eichstaedt. 2020. Estimating geographic subjective well-being from Twitter: A comparison of dictionary and data-driven language methods. *Proceedings of the National Academy of Sciences* 117, 19 (2020), 10165–10171.
[26] Wolfram Kawohl and Carlos Nordt. 2020. COVID-19, unemployment, and suicide. *The Lancet Psychiatry* 7, 5 (2020), 389–390.
[27] Nancy Krieger. 2001. A glossary for social epidemiology. *Journal of Epidemiology & Community Health* 55, 10 (2001), 693–700.
[28] Emily Leslie and Riley Wilson. 2020. Sheltering in place and domestic violence: Evidence from calls for service during COVID-19. *Journal of Public Economics* 189 (2020), 104241.
[29] Casey Littlejohn. 2019. Doing Harm: The Truth About How Bad Medicine and Lazy Science leave Women Dismissed, Misdiagnosed, and Sick, and editor of Feministing. com, explains the evolution of the current medical system in America; starting in the 18th century the existing medical community was overrun by European ideas, and female caretakers were eradicated (5-11). Without women working in the medical field, they were underrepresented and lacked a voice in their treatment. (2019).
[30] Joshua C Magee, Louisa Bigelow, Samantha DeHaan, and Brian S Mustanski. 2012. Sexual health information seeking online: a mixed-methods study among lesbian, gay, bisexual, and transgender young people. *Health Education & Behavior* 39, 3 (2012), 276–289.
[31] Christopher D Manning, Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*. 55–60.
[32] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. 3111–3119.
[33] Carmen Moreno, Til Wykes, Silvana Galderisi, Merete Nordentoft, Nicolas Crossley, Nev Jones, Mary Cannon, Christoph U Correll, Louise Byrne, Sarah Carr, et al. 2020. How mental health care should change as a consequence of the COVID-19 pandemic. *The Lancet Psychiatry* (2020).
[34] Nancy Morrow-Howell, Natalie Galucia, and Emma Swinford. 2020. Recovering from the COVID-19 pandemic: a focus on older adults. *Journal of aging & social policy* 32, 4-5 (2020), 526–535.
[35] Karsten Müller and Carlo Schwarz. 2020. From hashtag to hate crime: Twitter and anti-minority sentiment. *Available at SSRN 3149103* (2020).
[36] Marcia Mundt, Karen Ross, and Charla M Burnett. 2018. Scaling social movements through social media: The case of Black Lives Matter. *Social Media+ Society* 4, 4 (2018), 2056305118807911.
[37] Adrienne Nolan-Smith. 2021. MAYA DUSENBERY ON HOW MEDICAL GASLIGHTING SIDELINES WOMEN'S HEALTH PROBLEMS. https://getwellbe.com/medical-gaslighting-maya-dusenbery/.
[38] Abigail Oakley. 2016. Disturbing hegemonic discourse: Nonbinary gender and sexual orientation labeling on Tumblr. *Social Media+ Society* 2, 3 (2016), 2056305116664217.
[39] 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)*. 1532–1543.
[40] Joe Pinsker. 2020. THE PANDEMIC WILL CLEAVE AMERICA IN TWO. https://www.theatlantic.com/family/archive/2020/04/two-pandemics-us-coronavirus-inequality/609622/.
[41] Vincent J Roscigno, Diana L Karafin, and Griff Tester. 2009. The complexities and processes of racial housing discrimination. *Social Problems* 56, 1 (2009), 49–69.
[42] Koustuv Saha, Larry Chan, Kaya De Barbaro, Gregory D Abowd, and Munmun De Choudhury. 2017. Inferring mood instability on social media by leveraging ecological momentary assessments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–27.
[43] Koustuv Saha and Munmun De Choudhury. 2017. Modeling stress with social media around incidents of gun violence on college campuses. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–27.

[44] Koustuv Saha, Sang Chan Kim, Manikanta D Reddy, Albert J Carter, Eva Sharma, Oliver L Haimson, and Munmun De Choudhury. 2019. The language of LGBTQ+ minority stress experiences on social media. *Proceedings of the ACM on human-computer interaction* 3, CSCW (2019), 1–22.

[45] Koustuv Saha, Ingmar Weber, and Munmun De Choudhury. 2018. A social media based examination of the effects of counseling recommendations after student deaths on college campuses. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 12.

[46] H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Lukasz Dziurzynski, Stephanie M Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin EP Seligman, et al. 2013. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one* 8, 9 (2013), e73791.

[47] STAT. 2021. Luck is not a strategy: The world needs to start preparing now for the next pandemic. https://www.statnews.com/2021/05/18/luck-is-not-a-strategy-the-world-needs-to-start-preparing-now-for-the-next-pandemic/.

[48] Zsófia Szlamka, Márta Kiss, Sámuel Bernáth, Péter Kámán, Amina Lubani, Orsolya Karner, and Zsolt Demetrovics. 2021. Mental health support in the time of crisis: are we prepared? Experiences with the CoViD-19 counselling programme in Hungary. *Frontiers in psychiatry* 12 (2021), 792.

[49] Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology* 29, 1 (2010), 24–54.

[50] Riina Vuorikari, Anca Velicu, Stephane Chaudron, Romina Cachia, and Rosanna Di Gioia. 2020. *How Families Handled Emergency Remote Schooling during the COVID-19 Lockdown in Spring 2020: Summary of Key Findings from Families with Children in 11 European Countries. JRC Science for Policy Report.* ERIC.

[51] WIRED. 2020. Singapore Was Ready for Covid-19—Other Countries, Take Note. https://www.wired.com/story/singapore-was-ready-for-covid-19-other-countries-take-note/.

[52] Queenie Wong. 2021. Twitter, Facebook and others are failing to stop anti-Asian hate. https://www.cnet.com/news/twitter-facebook-and-others-are-failing-to-stop-anti-asian-hate/.

[53] Daniel Wood and Maria Godoy. 2020. What Do Coronavirus Racial Disparities Look Like State By State? https://www.npr.org/sections/health-shots/2020/05/30/865413079/what-do-coronavirus-racial-disparities-look-like-state-by-state.

[54] Xizhu Xiao, Porismita Borah, and Yan Su. 2021. The dangers of blind trust: Examining the interplay among social media news use, misinformation identification, and news trust on conspiracy beliefs. *Public Understanding of Science* (2021), 0963662521998025.

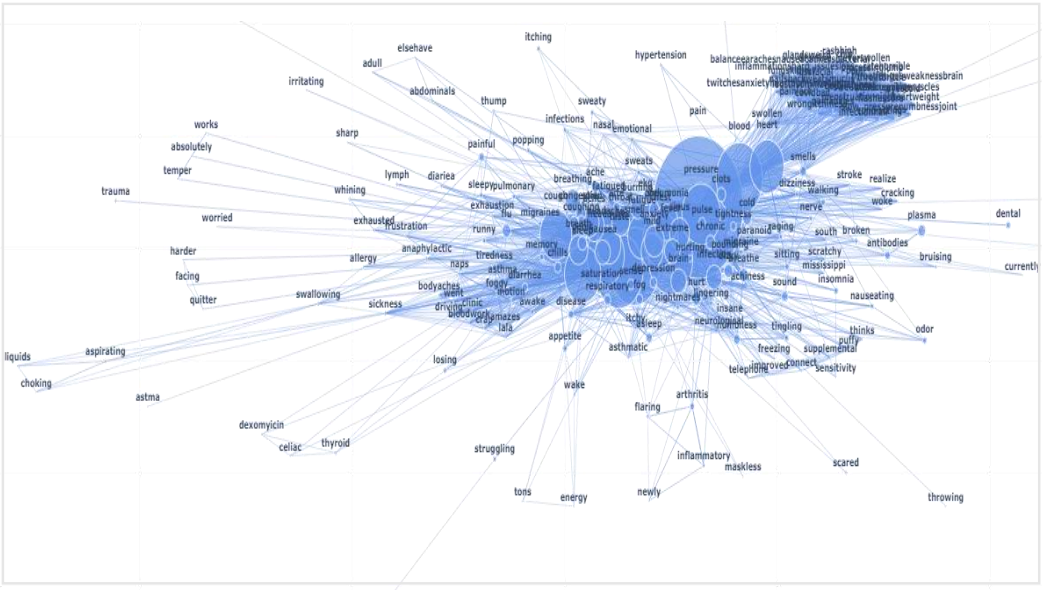


Fig. 6. A network graph showing all (264) COVID-19 Long-Hauler symptoms extracted from social media (Facebook) group's posts and comments.

7 APPENDICES