Doctors vs. Nurses: Understanding the Great Divide in Vaccine Hesitancy among Healthcare Workers

Sajid Hussain Rafi Ahamed
Goergen Institute for Data Science
University of Rochester
Rochester, NY, USA
srafiaha@ur.rochester.edu,

Shahid Shakil
Goergen Institute for Data Science
University of Rochester
Rochester, NY, USA
sshakil2@ur.rochester.edu

Hanjia Lyu
Department of Computer Science
University of Rochester
Rochester, NY, USA
hlyu5@ur.rochester.edu

Xinping Zhang
Department of Orthopaedics & Center for Musculoskeletal Research
University of Rochester Medical Center
Rochester, NY, USA
xinping_zhang@urmc.rochester.edu

Jiebo Luo
Department of Computer Science
University of Rochester
Rochester, NY, USA
jluo@cs.rochester.edu

Abstract—Healthcare workers such as doctors and nurses are expected to be trustworthy and creditable sources of vaccine-related information. Their opinions toward the COVID-19 vaccines may influence the vaccine uptake among the general population. However, vaccine hesitancy is still an important issue even among the healthcare workers. Therefore, it is critical to understand their opinions to help reduce the level of vaccine hesitancy. There have been studies examining healthcare workers' viewpoints on COVID-19 vaccines using questionnaires. Reportedly, a considerably higher proportion of vaccine hesitancy is observed among nurses, compared to doctors. We intend to verify and study this phenomenon at a much larger scale and in fine grain using social media data, which has been effectively and efficiently leveraged by researchers to address real-world issues during the COVID-19 pandemic. More specifically, we use a keyword search to identify healthcare workers and further classify them into doctors and nurses from the profile descriptions of the corresponding Twitter users. Moreover, we apply a transformer-based language model to remove irrelevant tweets. Sentiment analysis and topic modeling are employed to analyze and compare the corresponding Twitter users. In addition, we apply a key-term-based language model to remove irrelevant tweets. Sentiment analysis and topic modeling are employed to analyze and compare the corresponding Twitter users. Moreover, we apply a transformer-based language model to remove irrelevant tweets. Sentiment analysis and topic modeling are employed to analyze and compare the corresponding Twitter users.

Index Terms—COVID-19, vaccine hesitancy, healthcare worker, sentiment analysis, topic modeling

I. INTRODUCTION

COVID-19, a novel coronavirus, was identified in December 2019 as a potential cause of respiratory infections such as cough, flu, and pneumonia. The WHO (World Health Organization) Emergency Committee declared a global pandemic on March 11, 2020, due to the rapid spread of the virus worldwide [1]. As of August 29, 2022, there have been 602,479,782 infections and 6,480,637 deaths in total around the globe. Vaccine hesitancy indicates that vaccine availability does not guarantee adequate vaccination coverage. Previous research has shown that vaccine compliance is varied [2]. Socioeconomically disadvantaged groups are more likely to hold polarized opinions on COVID-19 vaccines, either pro-vaccine or anti-vaccine [2]. Health communications on social platforms about the advantage and safety of getting vaccinated does not necessarily increase vaccine uptake due to the negative impact of online social endorsement [4]. There are both challenges and opportunities in the process of designing effective communication and immunization programs [5]. Effective vaccinations against COVID-19 would need massive public education campaigns about vaccine safety and efficacy [6]. Medical professionals are among the group of front-line healthcare providers who are likely to be exposed to COVID-19 patients. Trust in healthcare workers is associated with increased vaccination intention [7]. Their opinions will be entrusted with providing vaccine recommendations and counseling vaccine-hesitant patients. Hence, research on vaccine hesitancy among healthcare workers is warranted. Multiple questionnaire-based studies have been conducted to examine the perceptions of the COVID-19 vaccines and willingness in vaccination among healthcare workers [8]–[13]. Interestingly, although doctors and nurses are both healthcare workers, there is a considerable difference in the level of vaccine hesitancy regarding COVID-

1https://covid19.who.int/
2https://www.usnews.com/news/national-news/articles/2019-01-16/who-names-vaccine-hesitancy-as-top-world-threat-in-2019
19. A higher proportion of nurses are COVID-vaccine hesitant compared with doctors [14].

In the era of the COVID-19 pandemic, social media data has been leveraged by researchers for various purposes [15] such as analyzing public attitudes [2], [16], identifying infodemics [4], [17], assessing mental health [18], [19], and so on. To our best knowledge, there have not been any social media studies on healthcare workers’ opinions about COVID-19 vaccines. To fill this gap, our study focuses on using social media data to understand medical practitioners’ opinions toward COVID-19 vaccines, the reasons for their hesitancy if any, and how this may translate to the general public’s opinion, as these factors may influence the general public’s vaccine decisions.

Specifically, this study uses the keyword search in user’s profile description to identify whether or not a Twitter user is a healthcare worker. Based on the keywords, the detected medical professionals are further divided into doctors, nurses, and others. We focus on comparing the sentiments and themes of the tweets posted by doctors and nurses. We find that doctors are more concerned with the ineffectiveness of the vaccines on the new variants while nurses pay more attention to the risk of vaccine side effects on children.

II. RELATED WORK

Vaccine hesitancy, defined as a reluctance to vaccinate despite the availability of vaccinations, is one of WHO’s top 10 global health hazards for 2019. Since then, several studies have analyzed the reasons for the public hesitancy toward various vaccinations.

Khubchandani et al. [3] launched a survey to assess the COVID-19 vaccine hesitancy in the community-based sample American population. The survey included 1,878 participants, most of whom were women, Whites, non-Hispanic, married, working full-time, and had a bachelor’s degree or above. In the study population, the likelihood of receiving a COVID-19 vaccine was: very likely (52%), somewhat likely (27%), not likely (15%), and not (7%), with individuals with lower education, income, or perceived risk of infection being more likely to report that they were not likely to receive the vaccine. Lyu et al. [2] proposed a human-guided machine learning framework to evaluate over six million tweets to categorize public opinion on COVID-19 vaccination and found that political affiliation and COVID-related experience are correlated with vaccination intention.

There are a few studies on understanding opinions toward COVID-19 vaccination among particular groups of population. For instance, Lucia et al. [6] found that students who were willing to receive the vaccination right away were more inclined to believe public health officials, had fewer concerns about adverse effects, and supported vaccine requirements. They highlighted the need for an educational curriculum to improve students’ awareness of COVID-19 vaccines and educate them on how to give vaccination advice. In terms of the vaccine hesitancy among medical professionals, Adejumo et al. [8] found that educational interventions to improve healthcare workers’ perceptions and attitudes toward the COVID-19 vaccine are also necessary. There are several common reasons for not getting vaccinated. Healthcare workers may have doubt on the efficacy of the vaccine, distrust of its content, and fear of side effects [11]. Building COVID-19 vaccine confidence may increase vaccine acceptance among them [9]. Our study intends to investigate this question at a larger scale and in fine grain with social media data.

III. MATERIAL AND METHOD

In this section, we describe how we collect Twitter data, identify medical professionals, and remove irrelevant data that do no disclose users’ opinions with respect to COVID-vaccine uptake. We also discuss how we estimate and analyze the sentiments and themes in the online discussion of doctors and nurses.

A. Data Collection

Following Lyu et al. [2], we use the Twitter API - Tweepy, to collect the publicly available tweets that contain particular keywords from September to November, 2021. The same keyword list of Lyu et al. [2] is used which is composed of “vaccine”, “COVID-19 vaccine”, “COVID vaccine”, “COVID19 vaccine”, “vaccinated”, “immunization”, “covidvaccine”, “#vaccine”, “covid19 vaccine”, “vinçine”, “vaccine”, “antivax”, and “anti-vax”. We remove the tweets that contain “vaccine” but are not related to COVID-19. For example, these tweets may discuss flu shots or polio vaccines.

B. Medical Professional Inference

Twitter users can disclose their professions in profile descriptions. For example, a Twitter user may write "Wife Mother Retired Nursing"

We use the keyword search to identify whether or not a user is a medical professional. First, we manually read the profile descriptions of 9,325 unique users and construct the keyword list related to the medical professions. We find 148 prominent keywords (Fig. 1) that identify the user’s profession. If the profile description does not contain any of the keywords, it is labeled as non-medical professional. Some example keywords are “doctor”, “nurse”, “icu”, “rn”, “pathologist”, “bsn”, “anthropology”, “epidemiology”, “medicine”, “emt”, “virologist”, “nephrologist”, “pharmacist”, “cardiology”, “surgeon”, “pharmd”, “psychiatry”, “therapist”, etc. We apply the keyword search to the remaining non-labeled data. We test our keyword search method on the labeled dataset. Table I shows the classification performance. In the sample of 152,213 unique users, 3,904 medical professionals are observed, accounting for about 3% of the sample size. This result is very much in line with the Census Bureau’s 2019 American
Community Survey\(^4\), where about 9.8 million workers are employed as health care technicians and practitioners, including physicians, surgeons, and registered nurses, which accounts for 4% of adults.

### TABLE I: Performance of the medical professional inference.

| Metrics     | Score |
|-------------|-------|
| F1 Score    | 90.7  |
| Precision   | 88.6  |
| Recall      | 92.9  |

Using these 148 keywords, we further classify the identified medical professionals into (1) doctors, (2) nurses, and (3) other medical professionals.

### C. Irrelevant Data Removal

Our goal is to have a better understanding of how medical practitioners think about COVID-19 vaccines. However, similar to Lyu et al. [2], not every tweet we collect reveals users’ opinions. They could simply be notifications related to vaccines. For example,\(^5\)

"Students should learn from home if the class does not have at least a 70% vaccination rate where the green passport restrictions are applied."

Therefore, we employ the XLNet model [20] from Lyu et al. [2], where they fine-tuned the model with 4,900 manually labeled data. The data were acquired in the same way as we do. They labeled them as relevant or irrelevant based on whether or not, by manual inspections, a researcher thinks the tweet expresses pro-vaccine, vaccine-hesitant, anti-vaccine opinions. If the tweet does not express any of the opinions toward COVID-19 vaccines, it will be labeled irrelevant. The model fined-tuned with these data achieves good performance on the relevant/irrelevant classification task (weighted precision/recall/F1-score = 0.70/0.63/0.63).

Consequently, we apply this model to our data and remove all the irrelevant tweets. It is noteworthy that before we apply this model, we adopt a tweet preprocessing pipeline from Baziotis et al. [21] which converts specific components such as handle names, URLs (Uniform Resource Locators), capitalized letters into special tokens. For example, if the original tweet is

"How the authoritarians among us will react to COVID vaccine mandates https://t.co/..."

After the preprocessing pipeline, it becomes

"how the authoritarians among us will react to < allcaps >covid< /allcaps > vaccine mandates < url >"

### D. Sentiment Analysis

In our study, we focus on investigating the sentiment toward the COVID-19 vaccines expressed by the healthcare workers. It is important to note that sentiment is not necessarily stance. According to Lyu et al. [2], the sentiment toward COVID-19 is positively correlated with the likelihood of holding pro-vaccine opinions. Therefore, we use sentiment as a proxy to estimate healthcare workers’ opinions.
For each tweet, we infer its sentiment using VADER (Valence Aware Dictionary for sEntiment Reasoning), a rule-based model for general sentiment analysis [22]. It calculates a normalized, weighted composite score to estimate the sentiment expressed by the tweet content. This score is between -1 (most extreme negative) and +1 (most extreme positive). If the score is greater than or equal to 0.05, the tweet is labeled positive. If the score is lower than or equal to -0.05, it is labeled negative. Otherwise, the tweet is labeled neutral.

E. Topic Modelling

Topic modeling is a method for unsupervised classification of documents, which finds a similar group of items (topics). A document can be part of multiple topics. We intend to analyze the relationship between the sentiments and the topics. We apply Latent Dirichlet Allocation (LDA) [23] to extract the topics from all relevant tweets. Similarly, we apply LDA to other four groups of tweets: (1) tweets posted by doctors with positive sentiments, (2) tweets posted by nurses with positive sentiments, (3) tweets posted by doctors with negative sentiments, and (4) tweets posted by nurses with negative sentiments. We choose the number of topics on the basis of both the coherence scores and manual inspections. The number of topics is set as five.

IV. RESULTS

Fig. 2 shows the sentiment distributions of the tweets posted by doctors and nurses. Among all tweets that are posted by doctors, 49.7% are positive, 30.4% are negative, 19.9% are neutral, while among all tweets that are posted by nurses, the proportions of positive, negative, and neutral classes are 39.1%, 37.5%, and 23.4%, respectively. After performing the Chi-squared test, we find the sentiment distributions of the tweets posted by doctors and nurses are significantly different ($p < .05$). Overall, tweets posted by doctors show more positivity about the COVID-19 vaccines.

Fig. 3 shows the word clouds of the most positive and negative tweets posted by doctors. We find that doctors talk about the effectiveness of the vaccines and urge people to get vaccinated in the positive tweets. In contrast, the pessimistic tweets reveal the risk involved in getting vaccinated.

Fig. 4 shows the word clouds of the most positive and negative tweets posted by nurses. In the positive tweets, the nurses express their feelings about caring for people, requesting people to get vaccinated and stay safe from the virus. They are cautious about the risk and illnesses that might be the side effects of vaccines, which is depicted in the negative tweets.

Fig. 5 shows the top 10 keywords in each topic extracted by LDA on all tweets. There are a variety of topics including effectiveness (Topic 0), risk for children (Topic 1), side effects (Topic 2), different vaccines (Topic 3), and anti-vaccine sentiment (Topic 4).

Fig. 6 shows the top 10 keywords of each topic of the positive/negative tweets posted by doctors or nurses. We assign a concise topic label to each of them to summarize the themes in Table II.

By analyzing the topics of the positive tweets, we observe that when talking positively about the COVID-19 vaccines, the focuses of doctors and nurses are similar. They both mention the safety and effectiveness, collective responsibility (i.e., the purpose of protecting others such as vulnerable children and patients). They discuss the immunity obtained by getting vaccinated. Further, doctors mention that about 95% of COVID patients getting admitted to the hospitals are unvaccinated, urging the public to get both vaccinations and the booster. The nurses are more concerned with the safety of the children, point out that the requirement of masks will not be there if vaccinated, and argue that vaccines are for obtaining immunization from COVID-19.

We also observe some common concerns of the doctors and nurses when they discuss COVID-19 in a negative way. The risk in getting vaccinated, mandates, and long-term illness are the most frequent topics among the negative tweets posted by
TABLE II: Topics of positive/negative tweets posted by doctors or nurses.

| Positive tweets of doctors | Negative tweets of doctors | Positive tweets of nurses | Negative tweets of nurses |
|---------------------------|----------------------------|--------------------------|--------------------------|
| Protect family and children | Ineffectiveness of vaccines on the new variants | Protect patients | Risk |
| Booster | Right to not get vaccinated | Mask requirement | Impact on children |
| Immunity | Long-term illness | Immunity | Mandate |
| Safety/effectiveness | Death/risk | Safety/effectiveness | Long-term illness |
| Science | Adverse affects of vaccines | Booster | Government regulations |

V. DISCUSSION AND CONCLUSION

In this study, we use the keyword search to identify medical professionals on Twitter and collect their discussion about the COVID-19 vaccines. We use the XLNet model to retain the tweets that reveal their opinions. We measure the sentiment of the tweets posted by doctors and nurses using VADER to estimate the opinions and further conduct a thematic analysis.

We find that doctors are overall more positive toward the COVID-19 vaccines than nurses. This is consistent with the findings of Browne et al. [14] where they found nurses are more vaccine-hesitant. There are both positive and negative tweets posted by the doctors and nurses. One potential reason that doctors are more positive could be that they are trained to take the trade-off when the benefits outweigh the risks, while the nurses do not and many even decide to quit their jobs to avoid vaccination.7

6https://www.cdc.gov/coronavirus/2019-ncov/vaccines/vaccine-benefits.html
7https://www.bloomberg.com/news/articles/2021-08-26/covid-vaccine-mandates-drive-some-nurses-to-leave-america-s-hospitals
and nurses focus on different aspects about the COVID-19 vaccines when discussing it negatively, although there are several common concerns such as the risks in getting vaccinated, mandates and long-term illness. This echoes the findings of previous studies [8], [24].

Some doctors feel their right to reject the vaccine is being violated, and they are being forced to take vaccines. They are skeptical about the efficacy of the vaccines on future variants and demand that more tests or trials to be done on the vaccines so that they do not lead to long-term illnesses. However, the nurses discuss more about their children’s health and argue that vaccines should not be mandated on children. Nurses are also worried about the long-term effect of vaccines. This suggests that the various focuses might be the reason that the levels of vaccine hesitancy of the doctors and nurses are different.

To increase vaccine uptake among healthcare workers, more customized strategies should be deployed. For example, more experiments results showing the effectiveness of vaccines or boosters on the new variants should be communicated with doctors. Experiment results of vaccination on children should be highlighted when communicating with nurses.

Our study has several limitations. First, we only focus on one social platform. Future work can analyze the social media posts across various platforms to obtain a more comprehensive understanding. Second, we use the keyword search method to detect medical professionals. This works well on Twitter, however, other methods may need to be developed on other platforms. Third, our work can be extended to examining the opinions of medical students.

VI. ACKNOWLEDGMENTS

This research was supported in part by the Goergen Institute for Data Science.

REFERENCES

[1] D. Cucinotta and M. Vanelli, “Who declares covid-19 a pandemic,” Acta Bio Medica: Atenei Parmensis, vol. 91, no. 1, p. 157, 2020.

[2] H. Lyu, J. Wang, W. Wu, V. Duong, X. Zhang, T. D. Dye, and J. Luo, “Social media study of public opinions on potential covid-19 vaccines: informing dissent, disparities, and dissemination,” Intelligent medicine, vol. 2, no. 01, pp. 1–12, 2022.

[3] J. Kubchandani, S. Sharma, J. H. Price, M. J. Wiblishauser, M. Sharma, and F. J. Webb, “Covid-19 vaccination hesitancy in the united states: a rapid national assessment,” Journal of Community Health, vol. 46, no. 2, pp. 270–277, 2021.

[4] H. Lyu, Z. Zheng, and J. Luo, “Misinformation versus facts: Understanding the influence of news regarding covid-19 vaccines on vaccine uptake,” Health Data Science, vol. 2022, 2022.

[5] W. Wu, H. Lyu, and J. Luo, “Characterizing discourse about covid-19 vaccines: A reddit version of the pandemic story,” Health Data Science, vol. 2021, 2021.

[6] V. C. Lucia, A. Kekela, and N. M. Afonso, “Covid-19 vaccine hesitancy among medical students,” Journal of Public Health, vol. 43, no. 3, pp. 445–449, 2021.

[7] A. De Figueiredo, C. Simas, E. Karafillakis, P. Paterson, and H. J. Larson, “Mapping global trends in vaccine confidence and investigating barriers to vaccine uptake: a large-scale retrospective temporal modelling study,” The Lancet, vol. 396, no. 10255, pp. 898–908, 2020.

[8] O. A. Adejumo, O. A. Ogundele, C. R. Madubuko, R. O. Oluwafemi, O. C. Okoye, K. C. Okonkwo, S. S. Owolade, O. A. Junaid, O. M. Lawal, A. C. Enikomohem et al., “Perceptions of the covid-19 vaccine and willingness to receive vaccination among health workers in nigeria,” Osong Public Health and Research Perspectives, vol. 12, no. 4, p. 236, 2021.