Detecting Conspiracy Theory Against COVID-19 Vaccines

1st Md Hasibul Amin  
Department of Computer Science  
University Of Houston  
Houston, United States  
mamin3@uh.edu

2nd Harika Madanu  
Department of Computer Science  
University Of Houston  
Houston, United States  
hmadanu@uh.edu

3rd Sahithi Lavu  
Department of Computer Science  
University Of Houston  
Houston, United States  
slavu@uh.edu

4th Hadi Mansourifar  
Department of Computer Science  
University Of Houston  
Houston, United States  
hmansourifar@uh.edu

5th Dana Alsagheer  
Department of Computer Science  
University Of Houston  
Houston, United States  
dralsagh@CougarNet.uh.edu

6th Weidong Shi  
Department of Computer Science  
University Of Houston  
Houston, United States  
wshi3@central.uh.edu

Abstract—Since the beginning of the vaccination trial, social media has been flooded with anti-vaccination comments and conspiracy beliefs. As the day passes, the number of COVID-19 cases increases, and online platforms and a few news portals entertain sharing different conspiracy theories. The most popular conspiracy belief was the link between the 5G network spreading COVID-19 and the Chinese government spreading the virus as a bioweapon, which initially created racial hatred. Although some disbelief has less impact on society, others create massive destruction. For example, the 5G conspiracy led to the burn of the 5G Tower, and belief in the Chinese bioweapon story promoted an attack on the Asian-Americans. Another popular conspiracy belief was that Bill Gates spread this Coronavirus disease (COVID-19) by launching a mass vaccination program to track everyone. This Conspiracy belief creates distrust issues among laypeople and creates vaccine hesitancy. This study aims to discover the conspiracy theory against the vaccine on social platforms. We performed a sentiment analysis on the 598 unique sample comments related to COVID-19 vaccines. We used two different models, BERT and Perspective API, to find out the sentiment and toxicity of the sentence toward the COVID-19 vaccine.

Index Terms—COVID-19, Conspiracy against COVID-19 Vaccination, Sentiment Analysis, Conspiracy Theory, BERT, COVID-19, Google Perspective

I. INTRODUCTION

Statistics show that COVID-19 is the most rapidly spreading and deadliest virus ever. Since the first COVID-19 case arose in December 2019 in China, 613.4 million people got affected, and more than 6.5 million people have died worldwide, according to WHO as of Sep 2022 [31]. Conspiracy beliefs and misinformation about this make this COVID-19 pandemic worst [1] [30]. Initially, contradictory statements about wearing the mask and lack of knowledge about the virus between researchers and the government helped to spread misinformation about the COVID-19 virus [29] [23].

Conspiracy theories have a destructive effect on society. It demeans the people’s interest and creates distrust of authentic information [24] [34] [15]. Generally, rumors spread faster than original news. For many years, people believed the 9/11 attack was an inside job [24] [11] [44]. Global digitization and the sudden social media boom open various doors for users to easily share false and misleading information [38] [27].

Furthermore, due to a lack of information about the effect of the COVID-19 vaccine and general health-related knowledge among people, social media users are promoting false stories about the vaccine’s side effects. A study in Italy found that over 2000 online articles, fake news tended to be shared a million times more [26] [9]. On the other hand, few individuals and politicians make this an issue to weaken the ruling party and achieve their personal agenda [30].

While researchers said that only the mass vaccination program could overcome this pandemic, this fake and false information creates distrust among the people and leads to vaccine hesitancy [20] [29] [41]. Also, underlying fear about vaccine side effects and instant information about it create vaccine reluctance among the people [16] [6].

Bill Gates and the 5G conspiracy show the difference in how social media users react and use the conspiracy for different purposes. In South Africa, the 5G controversy attracted minimal interest with debates on the cause and consequence of the virus and did not bother the government [29]. At the same time, it was a widespread belief that Bill Gates was behind all of this by launching a vast vaccination program [1] [14].

Global politicians also promote conspiracy beliefs and misinformation among the people. In addition, this conspiracy belief in the world people leads to distrust towards governments and World Health Organization (WHO), which causes lightness to government regulations about taking vaccines and wearing the mask, and maintaining Physical distance [43] [3] and thus helps to spread the COVID-19 pervasively. In the beginning, there was also a rumor that COVID-19 was created in LAB and spread by the Chinese government [38] [27].

This study investigates the conspiracy theory against this COVID-19 vaccine by analyzing comments on the online plat-
form. These could reduce the belief in rumors and entertain the vaccination program. We find out the sentiment and toxicity of the different comments using Google Perspective and the BERT model.

II. RELATED WORK

Conspiracy theory is quite common in North America. A survey found that about one-quarter to one-third of the population express conspiracy-related opinions [27]. In 2020 a survey in the U.S. indicated that about 5% think that COVID-19 is pre-planned, and 20% said it could be valid [37]. Rumors or Conspiracy theories are not always the case of pervasively false beliefs. People sometimes do it intentionally for geopolitical or satisfying individual motives [43] [4] [5]. For example, continuous denial of climate change because of global temperature is gameplay to delay the action against it by demeaning the value of scientific research [24].

Conspiracy has its way of supporting and strengthening its views. People start believing in something if they repeatedly see the same news on social media. People get influenced more by social media than by communications [10] [4]. Commercial Branding uses the same strategy to connect people. Conspiracy theories spread rapidly because of the psychological nature of people [5] [42]. Scientific news is less likely to be shared by people than conspiracy pages. Based on the research on 255,225 polarized users of scientific pages, 76.79% interacted, and among 790,899 conspiracy users, 91.53% interacted with conspiracy pages in terms of liking’ [4].

Besides social media, mainstream T.V. news and political agenda also have a significant influence on Conspiracy beliefs. A report shows that people on T.V. news who are liberal took this pandemic as a national threat [36] [5] [17]. Opposition for political and legal reasons also plays a vital role in vaccine hesitancy. Initially, few people believed that vaccines did spread from the lab, and it causes autism and infertility in teenage girls, and US Centers for Disease Control and Prevention fought to cover that information [42].

Text-Mining is a popular approach to detecting conspiracy theories [32]. Besides, there are numerous machine learning models for text mining, finding out the semantic content or tone of the sentence [27] [35] [45]. On the other hand, Deep neural network performance well in finding out the semantic content of a sentence [21]. BERT and Perspective API also determines the sentence’s sentiment, toxicity, and severity [39] [46] [18] [8] [40]. Biddlestone et al., 2020 use Confirmatory Factor analysis (CFA) and Structural equation modeling (SEM) to find out conspiracy belief and intention scale from a set of data collected from different user groups based on age, gender, personality, and education level [5] [13].

Pummerer et al., 2022 randomly survey people to find out who trust and support the government, maintain physical distance, believe in political COVID-19 conspiracy and have a conspiracy mentality [44]. Gagliardone et al., 2021 performed common word matching in twitter dataset. They find out the most popular keyword, phrases, and hashtags from the Dataset with the help of researchers [14].

Apart from harmful hate speech towards vaccines is also very common on social media. In Nigeria, political parties use these global controversies as a weapon to weaken the interim government [42] [20]. On the other hand, the belief that 5G and COVID-19 are linked leads to the burnout of the Cell phone tower in Europe [1]. Sherief et al. find out hate speech from Twitter data using Perspective API. They find the sentence’s toxicity and attack the commenter by analyzing the score range from 0-1 [12]. Melton et al., 2021 perform lexical sentiment analysis and topics modeling using Latent Dirichlet Allocation on vaccine-related comments on social media [28].

III. DATA

We initially collected 950 users’ comments from various online news portals and their Facebook pages manually. Later, we remove the duplicate comments with an almost similar word. We then manually read and labeled all the comments into two categories, 0 (No) and 1 (Yes). Where 0 means the comment is neutral or in favor of the COVID-19 vaccine, and one means it is against the COVID-19 vaccine. We show some sample data in Table I.

Then we clean the data. We removed the comments not in English as our primary focus on English text and finally kept 598 sample comments. We then pre-processed the previously labeled data by removing noise and stop words and converting all the tweets to lower cases.

This Dataset only includes English comments and is mainly from the North American zone, as our primary focus is to discover conspiracy theories against the COVID-19 vaccine in the United States. All these comments are public and do not include personal information, for example, name, location, or gender. Therefore, it does not go against the privacy act. We have only focused on COVID-19 vaccine-related posts and comments. The data and code is available on GitHub1.

TABLE I: SAMPLE DATA OF COVID-19 VACCINES TWEETS.

| Comments                                                | Conspiracy | Label |
|---------------------------------------------------------|------------|-------|
| After getting vaccine you catch heart diseases          | Yes        | 1     |
| Fully vaccination can reduce death rate for COVID-19    | No         | 0     |
| After the second dose of the Moderna vaccine, people get old | Yes       | 1     |
| Vaccination can have an impact on gender change         | Yes        | 1     |
| After getting one dose of the J & J vaccine             | No         | 0     |
| thank you, governor Newsom, for implementing mandates to keep our students and school staff as safe as possible amidst this newest COVID-19 surge. | No | 0 |

IV. METHODOLOGIES

Detecting conspiracy theories is a complex problem to solve. Finding conspiracy requires identifying the sentence

1https://github.com/AminHasibul/ConspiracyAgainstCovidVaccines
structure and its sentiment. As the language and structure of the sentences are very discrete in a different language, it is hard to detect and tokenize the word in the correct format [7]. Also, current social media users use emojis, unique jargon, and symbols to express their opinion, which is still a complex task to solve for text classification [32] [46] [35] [22].

As we had already cleaned and pre-processed our sample in the dataset preparation stage, we first counted the bag of words and word distribution. Then we predict a few incomplete words frequently used in the data and alter them with the proper abbreviation, for example – (vac,vaccn,vcn to the vaccine; CVD, covd to Covid; DCRS, decrees to decrease, fantastic to fantastic; grt to great). We also make sure the data is labeled and distributed evenly. Figure 1 shows the most common word distribution used frequently in tweets or comments. Figure 2 shows the label distribution of the data, where we see the same amount of positive and negative comments.

After that, we performed sentiment analysis on our labeled data by training the data using two different models, BERT Base and Google Perspective API. In 2018, Devlin et al. proposed Google’s BERT natural language processing model. The BERT model is simple to implement and Powerful. BERT (Bidirectional Encoder Representations from Transformers) is a state-of-the-art machine learning model for NLP tasks [8]. It shows excellent performance in multiple NLP tasks, and pre-training with fine-tuning has become a commonly used method in NLP tasks [33] [45] [21].

BERT model is a multi-layer bidirectional transformer encoder Architecture that completely changed the previous methodologies of pre-training generated word vectors and downstream NLP tasks. The BERT model can take both a single sentence and a pair of sentences as input parameters [8]. It uses Word Piece embedding with a 30000 token vocabulary [18]. BERT trains the sentence-level vector and extracts more information from the context [8] [25].

Similarly, Perspective API applies machine learning techniques to identify abusive comments. In addition to the flagship Toxicity attribute, the Perspective model finds the threat, identity attacks, profanity, sexually explicit, and insult. To detect inappropriate comments, Perspective employs machine
learning algorithms [18] [35]. The algorithms analyze the scores of a phrase based on how influential it is in a conversation. Developers and publishers can use the score to provide feedback to commenters and assist moderators in analyzing comments more quickly [19].

V. EXPERIMENT, RESULT & EVALUATION

We split data into training and validation sets to validate the model performance. We train the model using BERT and Perspective. From Perspective API, we get the different scores (toxicity, severity, threat, attack, sexually explicit, profanity, insult). Table II shows some samples of Google Perspective score accounting on the comments.

| Comments                                    | Label | Toxicity | Severe | Identity | Attack | Insult | Profanity | Threat | Sexually | Explicit | Flirtation |
|---------------------------------------------|-------|----------|--------|----------|--------|--------|-----------|--------|----------|----------|-----------|
| A vaccine may temporarily affect your period | 1     | 0.0561   | 0.03674| 0.04487  | 0.09897| 0.04582| 0.07309   | 0.0368 | 0.17256  |
| Half of you are crazy. Why would anyone take the vaccine | 1     | 0.6758   | 0.35763| 0.23627  | 0.88076| 0.38093| 0.15138   | 0.0479 | 0.26543  |
| Mike Esher Like Colin Powell died. Fully vaccinated. My point exactly. | 1     | 0.0458   | 0.04646| 0.02733  | 0.03551| 0.01963| 0.25159   | 0.0178 | 0.2105   |
| I will be happy with my two Pfizer vaccines^* | 0     | 0.0721   | 0.087  | 0.09076  | 0.06088| 0.10754| 0.32974   | 0.1811 | 0.40821  |
| But your vaccine will protect you!          | 0     | 0.0196   | 0.0197 | 0.02333  | 0.05805| 0.0153 | 0.08437   | 0.0163 | 0.22616  |
| The vaccine is a blessing to us             | 0     | 0.0398   | 0.03101| 0.07103  | 0.04766| 0.05292| 0.09777   | 0.0649 | 0.23443  |

Then, we train using the BERT model. We used BERT-Base, Uncased, which is a 12-layer model and has 768-hidden layers, 12-heads, and 110M parameters. We then calculated the Accuracy, F1-score, Precision, and Recall for the newly trained model using Perspective API. Using a Logistic regression classifier, we get an Accuracy of 55%, a Recall and Precision score of 55%, and an F1 score of 53%. For the XGBoost classifier, we get an Accuracy, Recall, and Precision score of 65%, F1 score of 63%, and for the Gaussian Naïve Bayes classifier, we get an Accuracy, Recall, and Precision score of 75%, F1 score of 70% (Figure 4).

VI. COMPARISON

Figure 4 shows that Google Perspective with Gaussian Naïve Bayes classifier performance is (75%) better than others. BERT Model with Logistics Regression gives us an accuracy of 69%, and XGBoost accuracy is 66% which is close to L.R. But with Perspective Logistic Regression Performance is very poor, and Naïve Bayes with BERT model performs very poorly. Also, Google Perspective can detect which is abusive and which is not. We also notice an 8-9% increase in performance for both models by increasing the volume of data in the Dataset. Initially, we performed the same experiment on 400 data, and we got 61% for BERT Model with the L.R. classifier and 66% for Perspective with N.B. Classifier. So, we are assuming an increase in performance by adding more data to the Dataset and retraining the model.

VII. LIMITATION

We collected our sample data from North American-based users’ comments. The sample data is manually collected and not large enough to perform conspiracy theories for the entire world. So, it does not accurately reflect the thoughts of people from other regions. Also, we did not collect user information, so this dataset is not able to find out the age wise conspiracy belief. On the other hand, it is tough to detect which comments are valid and which are false in some cases.

VIII. CONCLUSION

Our study analyzed social media’s comments score and public sentiments related to COVID-19 vaccines. As most social media post comments have no structure and many use jargon to express their opinion, it is tough to classify those data [22] [2]. Our study found vaccine hesitancy among people by analyzing public sentiment and comments scores using BERT and Perspective API. Also, we found that comments in favor of vaccines are lower than the comments which are not in favor. Although our study gives an idea about public sentiment in North American regions, further study is needed to detect conspiracy and the user group for the other part of the world to get a more realistic picture of COVID-19 Vaccines.

ACKNOWLEDGMENT

The Authors would like to thank the COSC 6376 Cloud Computing Course Fall 21 – University of Houston students for collecting and sharing their data with us. It helps to conduct our research with a more extensive dataset.
PERFORMANCE COMPARISON

Fig. 3. Performance Comparison of Two Models.

Perspective Result

Accuracy
Precision
Recall
F1

BERT Model Result

Accuracy
Precision
Recall
F1

Fig. 4. Perspective Accuracy, F1 Score, Recall, and the Precision.

Fig. 5. BERT Model Accuracy, F1 Score, Recall, and Precision.

REFERENCES

[1] Wasim Ahmed, Josep Vidal-Alaball, Joseph Downing, Francesc López Seguí, et al. Covid-19 and the 5g conspiracy theory: social network analysis of twitter data. Journal of medical internet research, 22(5):e19458, 2020.

[2] Areej Al-Hassan and Hmood Al-Dossari. Detection of hate speech in social networks: a survey on multilingual corpus. In 6th International Conference on Computer Science and Information Technology, volume 10, pages 10–5121, 2019.

[3] Paul Bertin, Kenzo Nera, and Sylvain Delouvée. Conspiracy beliefs, rejection of vaccination, and support for hydroxychloroquine: A conceptual replication-extension in the covid-19 pandemic context. Frontiers in psychology, page 2471, 2020.

[4] Alessandro Bessi, Mauro Coletto, George Alexandru Davidescu, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi. Science vs conspiracy: Collective narratives in the age of misinformation. PloS one, 10(2):e0118093, 2015.

[5] Mikey Biddlestone, Ricky Green, and Karen M Douglas. Cultural orientation, power, belief in conspiracy theories, and intentions to reduce the spread of covid-19. British Journal of Social Psychology, 59(3):663–
