DIGITAL HEALTH

Understanding COVID-19 vaccine hesitancy of different regions in the post-epidemic era: A causality deep learning approach

Yang Liu[1] (ID), Chenxu Zhao[2] and Chengzhi Zhang[3]

Abstract

Objective: This paper aims to understand vaccine hesitancy in the post-epidemic era by analyzing texts related to vaccine reviews and public attitudes toward three prominent vaccine brands: Sinovac, AstraZeneca, and Pfizer, and exploring the relationship of vaccine hesitancy with the prevalence of epidemics in different regions.

Methods: We collected 165629 Twitter user comments associated with the vaccine brands. The comments were labeled based on willingness and attitude toward vaccination. We utilize a causality deep learning model, the Bert multi-channel convolutional neural network (BertMCNN), to predict users' willingness and attitude mutually.

Results: When applied to the provided dataset, the proposed BertMCNN model demonstrated superior performance to traditional machine learning algorithms and other deep learning models. It is worth noting that after March 2022, the public was more hesitant about the Sinovac vaccines.

Conclusions: This study reveals a connection between vaccine hesitancy and the prevalence of the epidemic in different regions. The analytical results obtained from this method can assist governmental health departments in making informed decisions regarding vaccination strategies.

Keywords

Vaccine hesitancy, the post-epidemic era, causality deep learning

Introduction

COVID-19, a highly devastating epidemic discovered in 2019, has led to significant disruptions in various sectors worldwide.[1] Several countries have been profoundly affected by the outbreak. Despite the development of COVID-19 vaccines in November 2020 by pharmaceutical companies like Pfizer,[2] research on the virus and its vaccination continues.[3] Unfortunately, some individuals skeptical of vaccines have raised concerns about potential side effects, contributing to vaccine hesitancy and the spread of conspiracy theories in the post-epidemic era.[4] Furthermore, researchers and public health experts have been working diligently to understand vaccine hesitancy and public attitudes or willingness in the post-epidemic era.[5]

Vaccine hesitancy refers to hesitancy about getting vaccinated,[6] which has been identified as a major threat to global health. To address this issue, various interventions have been implemented around the world to disseminate scientific knowledge to promote vaccination.[7] However, such attempts are often ineffective and may even have the

[1]School of Information Management, Wuhan University, Wuhan, China
[2]School of Computer Science, Wuhan University, Wuhan, China
[3]Department of Information Management, Nanjing University of Science & Technology, Nanjing, China

Corresponding author:
Yang Liu, School of Information Management, Wuhan University, Wuhan, China.
Email: yang.liu27@whu.edu.cn

opposite effect than expected.[8] In particular, people's reactions to vaccination campaigns in different places are negatively influenced by their attitudes and willingness.[9] Particularly in social media, negative public discussion can manifest as comments of anti-vaccination propaganda messages.[2] Indeed, anti-vaccine words have proliferated on social media. For example, after the introduction of COVID-19, 85% of Facebook comments on vaccines have been in anti-vaccine status, describing the severe side effects of vaccines and spreading conspiracy theories.[10]

The Internet has become the main source of almost all information, especially information on vaccination for most people.[11] This kind of information spreads quickly and many people express their opinions through social media platforms.[12] Thus, there is a need to use the vast wealth of data generated by social media, which detect concerns about the issue of vaccination in the population and control for changes in the users' behavior.[13] To do this effectively requires artificial intelligence technology to monitor epidemic situations.[14] Given the launch of the COVID-19 vaccine, the real-time use of artificial intelligence to support vaccination campaigns is more important than ever.[15] Health organizations are implementing social listening to identify and understand posts about vaccines on social media, which understand what topics are being discussed, whether information is being shared, and whether it is accurate.[16] Machine learning is already widely used with the COVID-19 vaccine. However, due to the limitations of machine learning in processing large amounts of textual data for feature extraction, deep learning can effectively improve feature extraction.[17] Moreover, this paper contributes to this growing literature, including using deep learning techniques to improve vaccination and protect society.

The main interest of this paper lies in developing a classification method for the users' willingness and attitudes that can detect indecision on social media in different regions. Unlike the traditional sentiment classification, this paper develops two categories of willingness and attitude. Therefore, we specifically address the problem of vaccine hesitancy following the classification of willingness and attitude. In addition to detecting the two labels, we also found that the attitude detection task can benefit from the willingness label as an additional input and vice versa. This finding supports the theory of causal inference,[18] which emphasizes causal. This paper utilizes a dependency tree-based graph convolutional network to construct the model, which can better capture the syntactic dependencies and improve the model performance. The experimental results showed a higher hesitation towards the Sinovac vaccine after March 2022. This result truly reflects the attitudes and willingness of different regions towards vaccines after the epidemic, fully allowing the government to understand the people's needs. Our discoveries aid policymakers and medical institutions in formulating more efficient approaches to tackle issues pertaining to the vaccination process.

To this end, we analyzed people's opinions on COVID-19 on Twitter, investigating public willingness and attitudes toward vaccines using deep learning. In summary, the main contributions of this paper are as follows: First, it employs deep learning to categorize vaccine hesitancy and utilizes a causal inference model to enhance the interpretability of the results. Second, the multi-channel model we propose analyzes tweets with multiple emotions to identify the reasons for public vaccine refusal, and the experimental results surpass those of the baseline model. Finally, the new insights gained from this research are crucial, aiding governments and vaccine companies understand the reasons for vaccine hesitancy and thereby gaining public trust. This contributes to the effective control of the post-pandemic era of the COVID-19 pandemic.

The remainder of the paper is structured as follows: In Section 2, we review pertinent related work. Section 3 furnishes a comprehensive description of our proposed approach. In Section 4, we showcase our experiment and discuss the results obtained. Section 5 delves into a discussion concerning the presented work. Lastly, Section 6 summarizes the work presented, outlines potential internal threats to validity, and discusses possible research extensions.

Literature review

Vaccine hesitancy in social media

Since the onset of the COVID-19 pandemic, numerous studies have employed sentiment analysis to investigate COVID-19 vaccination hesitancy. Researchers employ diverse techniques to scrutinize public sentiments surrounding COVID-19.[17] The general public utilizes Twitter datasets to express their opinions on the matter. Historically, prior literature studies have predominantly applied sentiment analysis in machine learning. In contemporary contexts, the primary objective of sentiment analysis is to comprehend individuals' attitudes toward the ongoing epidemic, particularly on social media.[2] A study conducted by Li et al. delved into the behaviors of Americans and Chinese on various social media platforms amid the COVID-19 pandemic.[19] The findings of their sentiment analysis revealed notable disparities in the attitudes between the two countries, with prevailing confidence among most individuals in controlling the virus's spread. Another investigation by Zhou et al. focused on extracting tweets from the pandemic period,[20] examining the emotional dynamics of Australians, and observing temporal shifts in their emotions. In a separate study, Yin et al. analyzed the sentiment of tweets related to COVID-19,[21] discovering a slight predominance of positive tweets over

negative ones. The analysis provided tweets emphasizing themes such as securing one's home and commemorating those who succumbed to COVID-19 in the post-epidemic era.

Table 1 presents a comprehensive comparative study on sentiment analysis related to COVID-19 vaccines. The juxtaposition of previous studies will facilitate the current research in exploring uncharted territories that were not addressed in earlier investigations. Additionally, it will serve as a valuable guide for forthcoming studies, providing insights into making unique contributions to the subject matter. This summary table encapsulates diverse studies employing distinct approaches, all conducted using Twitter datasets specifically focused on the study of COVID-19 vaccination hesitancy.

Machine learning in vaccine hesitancy

The COVID-19 pandemic underscores the imperative for advanced technology to respond to emergencies effectively. In tackling such scenarios, machine learning techniques emerge as valuable tools, offering robust support to public health authorities and complementing traditional approaches. For instance, Tavoschi et al., employed support vector machines to monitor public sentiment regarding vaccine uptake in Italy.[30] Moreover, social media traffic enhanced the performance of various models. Despite the utilization of machine learning in the studies, challenges such as poor classification effectiveness and limited generalization ability have prompted a shift toward deep learning as a viable solution. Hussain et al. applied a deep learning BERT model to analyze public sentiment toward COVID-19 in the UK and the USA.[31] Similarly, Devaraj et al. utilized a multivariate LSTM model for predicting cumulative confirmed cases and deaths in COVID-19 cases, achieving superior accuracy.[32] In another vein, Shastri et al.[33] conducted a comparative analysis of deep learning methods for predicting COVID-19 cases in the USA and India one month in advance.

The work above is confined to sentiment computation technology, where researchers have employed a singular computational approach.[4] We believe that comparing deep learning techniques will enhance the study of COVID-19 vaccine hesitancy in the post-pandemic era. We are convinced that employing multi-channel and multi-task modeling techniques can strengthen research on Covid-19 vaccine hesitancy. Therefore, another critical consideration is determining which methodological approaches are most effective for classifying COVID-19 vaccine hesitancy using post-pandemic datasets. Given the gaps in previous research on COVID-19 vaccine hesitancy, as indicated in Table 1, this study utilizes text mining, sentiment analysis of COVID-19 Twitter datasets, and machine learning techniques to understand public perceptions of COVID-19 vaccine hesitancy in the post-epidemic era. To achieve this goal, we designed, developed, and evaluated three different brands of Twitter COVID-19 classifiers, using causally informed deep learning models to construct optimal models for identifying public hesitancy.

Methods

Data collection

Data preprocessing. We gathered datasets daily from the public tweets stream using the Twitter API. We employed the API's search function with parameters such as 'q' for search words, 'lang' for language set to English ('en'), 'since', and 'until'. The search words included #vaccine or #covid19. #CovidVaccine, #AstraZeneca, #Sinovac, and #Pfizer. A notable challenge was the limitation imposed by Twitter, allowing researchers to access API search for only about 7–9 days. Consequently, we gathered COVID-19 vaccine-related tweets that included various predefined hashtags from January 1, 2022 to December 30, 2022. A total of 201960 tweets were collected during this period. Each day's live public tweets were saved as CSV (comma-separated value) files. Upon concluding the data collection, the day-to-day public tweets were aggregated into a comprehensive dataset.

Data preprocessing. The raw crawled tweets encompass diverse non-textual elements, including images, emojis, and symbols like #, $, etc. Moreover, retweets and bot-generated content significantly influence analysis outcomes as well. This poses a considerable challenge to our research, necessitating preprocessing of raw data to derive usable final text data.

Retweets and duplication. Retweets occur when users share others' tweets on their accounts, leading to potential data duplication. To address this, we utilized Python's 're' library to match the format of retweeted tweets, thereby removing duplicates from our collection.

Bot-generated contents. Given the advancements in artificial intelligence, Twitter has seen a surge in bot-generated tweets. These automated posts often fail to represent genuine user opinions, impacting our analysis. Identifying and excluding this text segment from our research was imperative. Many researchers have conducted extensive studies on bot detection, yielding various recognition algorithms. Leveraging the methods outlined, we identified and subsequently removed bot-generated tweets.

Emojis and punctuations. Non-textual symbols present in the data impact our analysis. Following duplicate removal, we employed Python's 're' library to match and

Table 1. A comprehensive summary of the results of available studies on the topic.

Paper	Data	Sentiment Techniques	Method	Conclusions
Amrita Mishra et al.,[22]	Daily tweets of vaccines Sputnik, Moderna Covaxin	TextBlob	Machine learning	The U.S. Moderna Vaccine is the most prominently discussed brand among Twitter users, garnering the highest favorability compared to Covaxin (India) and Sputnik V (Russia) vaccines. This research suggests a prevailing positive sentiment in tweets related to COVID-19, indicating a potentially higher acceptance rate for the COVID-19 vaccine than previous ones.
Joanne Chen et al.,[23]	Tweets maintained by Georgia State University's Panacea Lab	Textstem	Machine learning	According to this research, the increasing positive sentiments in the tweets about COVID-19 may imply that the vaccine has higher acceptance rates than previous vaccines.
Siru Liu et al.,[24]	Collected tweets using snscrape package in Python to collect tweet IDs and then used the tweepy package to collect the data	VADER	Pruned Exact Linear Time (PELT) algorithm	The study reveals that approximately 42.8% of tweets express positive sentiments, while 30.3% convey negative sentiments. Public opinion experienced an upswing following Pfizer's announcement of a 90% vaccine effectiveness.
Shamrat, et al.[25]	tweets using API search	TextBlob	KNN	Specifically, Pfizer exhibits a positive sentiment of 47.29%, Moderna follows closely at 46.16%, and AstraZeneca maintains a rate of 40.08%.
[21]Rashid Khan et al.[26]	Tweets collected by the Department of Computer Science, Abbottabad University	TextBlob	Random Forest, Gradient Boosting Machine, Extra Tree Classifier, Logistic Regression, and Support Vector Machine	In line with this research, incorporating TF-IDF features has notably enhanced the performance of the Gradient Boosting Machine, achieving an impressive 96% accuracy.
Ghulam Musa et al.[27]	Tweets using Twitter Scraper using Keywords 41,349 tweets using Twitter Scraper or Data from Kaggle	TextBlob	Count Vectorizer and TF-IDF using Logistic Regression, Multinomial Naïve Bayes, Decision Tree, Random Forest, Support Vector Machine and XGBoost	The peak accuracy is attained when employing the Support Vector Machine classifier with TF-IDF vectorization (93.15%) and Count Vectorizer (93.07%). Notably, TF-IDF accuracy is 10% more efficient than Count Vectorizer.

(continued)

Table 1. Continued.

Paper	Data	Sentiment Techniques	Method	Conclusions
Wang et al.[28]	Collected tweets using Tweeter API search	Spacy	Doc2Vec and K-means clustering	Drawing insights from this research, four crucial topics regarding the Covid-19 vaccine emerge for Australians: national vaccine rollout, the correlation between vaccines and mortality, vaccine approval processes, and vaccine hesitancy.
Qorib et al.[15]	Datasets were collected daily from the public tweets streaming using Twitter API.	Azure Machine Learning, VADER, and TextBlob	Random Forest, Logistics Regression, Decision Tree, Linear, and Naïve Bayes	The experimental results demonstrate that the combination of TextBlob, TF-IDF, and LinearSVC yields the optimal performance in classifying public sentiment into positive, neutral, or negative categories. The accuracy, precision, recall, and F1 score stand at 0.96752, 0.96921, 0.92807, and 0.94702, respectively.
Erika et al.,[29]	Retrieved data from PGP analysts	Comparing conversation four months before and after COVID-19 spread	Machine learning	Notably, vaccine opponents actively provoke resistance against the COVID-19 vaccine and foster mistrust in health authorities. Exposure to increased vaccine opposition may have a tangible impact on the overall health of a population.
This paper	Collected tweets using Tweeter API search	BERT	Dual-channel deep learning	This study reveals a connection between vaccine hesitancy and the prevalence of the epidemic in different regions. The analytical results obtained from this method can assist governmental health departments in making informed decisions regarding vaccination strategies.

eliminate symbols within tweets. Additionally, we utilized the emoji library to strip emojis from the tweet content.

Tokenization and stop words. Tokenization involves breaking down sentences into individual words, facilitating a more streamlined analysis. However, this process often generates common words like 'the,' 'is,' ''are,' etc., known as stop words, which need more contextual meaning. Removing these stop words aids in deciphering user intent and sentiment. To accomplish this, we leveraged this study's NLTK library for tokenization and stop word removal.

Annotation. We eliminated retweets and duplicate tweets to ensure data accuracy, focusing solely on English tweets. Figure 1 illustrates the complete process of data collection and annotation.
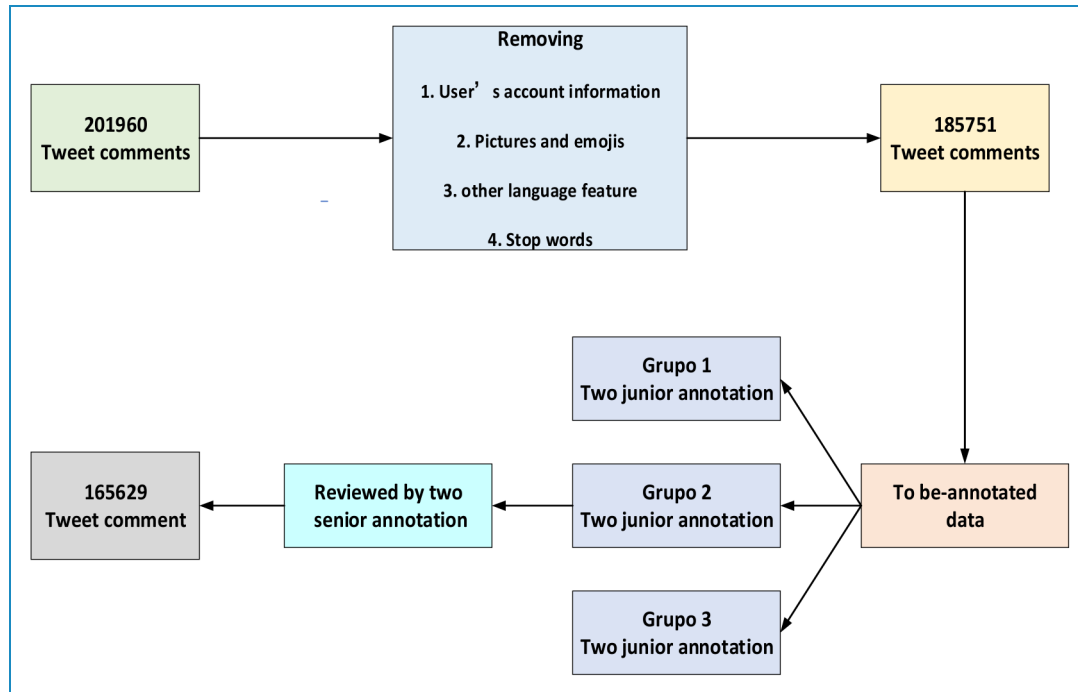
Figure 1. The process of data collection and annotation.

After removing retweets and duplicate tweets, there were 185751 tweets remaining. Following the annotation process and removal of controversial annotations, the final count came down to 165629 tweets. Controversial refers to the inability to label accurately or the labelers' opinions are inconsistent when labeling. Although the date of the user's account comments is valuable for analyzing topic timeliness, we currently do not require this information during the topic modeling stage. Therefore, we have excluded this data and retained only the annotation text. Nevertheless, it is worth noting that pictures and emojis can be valuable for sentiment analysis purposes.[34] We have focused solely on text data and left the multimodal data analysis for future research. Consequently, all pictures and emojis have been removed from the dataset. The primary objective of this paper is to analyze English comments; therefore, we have filtered out non-English characters, retaining only English comments for analysis.

As for the annotation process, it involved the participation of two medical experts and six graduate students of social science who formed the annotation team. The team of six students took on the role of annotators, responsible for labeling each Twitter comment, while the two experts served as team leaders, overseeing the entire process. To establish consistent and reliable annotations, a back-to-back labeling method was adopted. The six students were divided into three groups, and each tweet was annotated by two students simultaneously. Subsequently, the annotation team discussed any discrepancies, and the two experts made final decisions in cases of inconsistent labels.

Comments with contentious annotations were discarded to ensure data integrity. Table 2 shows a sample of the data after the processing was completed.

Among these tweets, each data point was labeled with the user's attitude towards the brand of the vaccine and their willingness to get vaccinated. In detail, attitude labels are divided into three categories: Positive, Neutral, and Negative, and willingness labels are divided into three categories: Yes, Not Sure, and No. After excluding the controversial comments, 165629 comments remained for further analysis. As a result, the annotator's agreement score was deemed 100%, indicating a high level of consistency and accuracy in the annotation process. Then we categorize the three brands. The quantities of the three brands are shown in Table 3. For example, the Pfizer vaccine is produced in the United States, so most people in the United States received this brand of vaccine. The distribution of each label for the three brands is shown in Figure 2, with the attitude label distributed more evenly and willingness dominated byNot Sure.

It is evident that the distribution of labels in the data related to *attitude* exhibits a fundamental balance. However, in the context of *willingness* distribution, theNo label stands significantly lower than the others. Consequently, during the model training process, we opted to oversample the data labeled as 'No' to a certain degree to uphold a balanced data distribution. Specifically, we employed Easy Data Augmentation (EDA) 34 techniques to enrich the "No" label dataset. This was achieved through various strategies, including random insertion (RI), random swap (RS), synonym replacement (SR),

and random deletion (RD). The label distribution after oversampling is shown on the right in Figure 2. After that, we divided the dataset into a training set, validation set and test set in the ratio of 60%, 20%, and 20%.

Table 2. Examples of the processed data.

Tweets	Brands	Willingness	Attitude
Is the AstraZeneca COVID vaccine effective?	AstraZeneca	Not sure	Neutral
Nice – Queen of the Vaccines on the AstraZeneca segment!	AstraZeneca	Not sure	Positive
Excellent news. Which vaccine? We had Pfizer.	Pfizer	Yes	Positive
it's the pfizer vaccine in me i'm sorry.	Pfizer	Yes	Negative
Sinovac is as good as water...it is a profit vaccine.	Sinovac	No	Negative

Table 3. Statistics on the number of datasets.

Country	Brand	Number of tweets
USA	Pfizer	59070
UK	AstraZeneca	48955
China	Sinovac	57604

Model

Taking inspiration from various text classification models like BertCNN, LSTM, and TextGCN, we hypothesize that GCN and LSTM have distinct abilities idn extracting information at different levels, while CNN excels in synthesizing diverse localized features. Consequently, based on these insights, we introduced a graph convolutional networks multi-channel model (GCNMC). Our hypothesis is further substantiated in the subsequent experimental section.

Figure 3 offers an overview of the whole model. Mathematically, we defined the standing detection classification task. Given a short text $Text = \{w_1, w_2, w_3, \ldots\ldots, w_{n-1}, w_n\}$, in addition to predicting the two labels willingness and attitude by *Text* alone, we also input both *Text* and *willingness (attitude)* to predict *attitude (willingness)* because there is a causal relationship between them. Specifically, we input one of the labels into the model immediately after *Text* and predicted the other label.

Embedding layer. We feed our context into bidirectional encoder representations from transformers (BERT)[35] embedding, whose output is a sequence of high dimensional vectors. These vectors reflect information from the whole context. Specifically, BERT has two types of scale, which has different numbers of parameters, $Bert_{base}$ and $Bert_{large}$. The size of the hidden dimension in $Bert_{base}$ is 768 while the size of the hidden dimension in $Bert_{large}$ is 1024. We adopted $Bert_{base}$ in our model. It has 12 layers and 12 attention heads and contains 110 M parameters.

Graph neural network layer. In our work, we built a graph based on dependency trees. Offered by Stanford CoreNLP[36] tools, we leveraged the dependency parser in the package to gain relationships between different words in a short text. Mathematically, based on the dependency tree, we constructed the graph $G = \{E, V, M\}$, where V
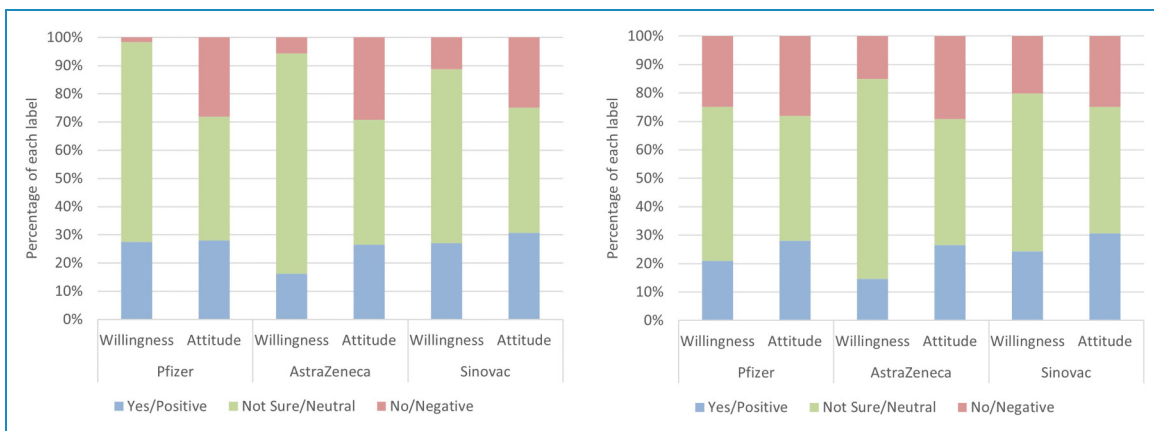

Figure 2. Distribution of each label; the left is before oversampling while the right is after oversampling.
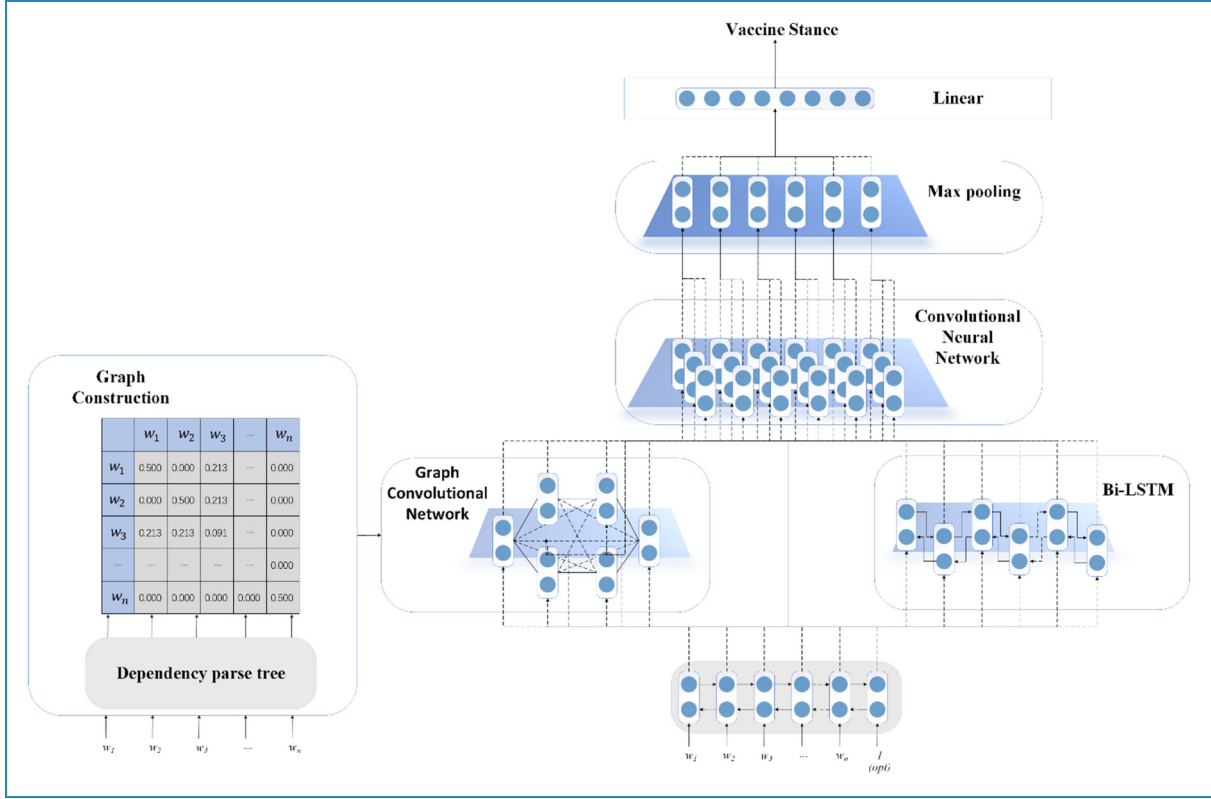
Figure 3. The structure diagram of our proposed model.

indicates the nodes set that contains all words of the whole text and E means the edge set that contains all edges between the different nodes or words. For each edge e, if the two words it connects have a relationship in the dependency tree, then we set it as 1 in the adjacent matrix M. It can be described as:

$$m_{ij} = \begin{cases} 1 \ if \ e_{ij} \in E, \\ 1 \ if \ i = j, \\ 0 \ otherwise \end{cases}$$

Then, we derived the weight matrix X from the adjacent matrix M. To be specific, like,[37] we normalized the adjacent matrix with its degree matrix D, and it can be described as follows:

$$X = D^{-\frac{1}{2}} M D^{-\frac{1}{2}}$$

where D can be calculated as:

$$D_{ii} = \sum_{j} m_{ij}$$

For the node features, we adopted the output of the Bert embedding as the node features. In detail, the feature matrix $F \in R^{l*dim}$ is the H mentioned in the last part, which will be updated by the Graph Neural Network with the edge weights matrix.

Multi-channel layer. To extract all kinds of features and information in the sequence of the text,[38] we designed this layer, which contains different types of components like the Graph Neural Network and Bidirectional Long Short-Term Memory (Bi-LSTM). There are two LSTMs that compute the given vectors from different directions. After that, we fed the graph and the node features into graph convolutional network. The node features updates with the convolution of the GNN and finally output the features with the same dimension of the input node features.

Mathematically, the graph $G \in R^{l \times l}$ is a normalized matrix that will not change with the computation of the GCN while the node features, that is the embedding vectors, $H \in R^{l \times dim}$, change. In fact, this feature matrix is calculated as:

$$O_i^{GCN} = ReLU(X O_{i-1}^{GCN} W_i + b_i)$$

where O denotes the output of the i GCN layer, X denotes the normalized edge weights matrix, W denotes the parameters of the i GCN layer and b denotes the bias of the i GCN layer.

We concatenated different feature matrices derived from the abovementioned channels to enrich the feature. In detail, we extracted dependency relationships feature and contextualized features through GNN and the Bi-LSTM,

respectively. We merged these two channels, which means we concatenated O and O. Additionally, to combine low level features, we also directly concatenated the Bert embedding with them as well. It can be described as follows:

$$F = Concat(O^{LSTM}, O^{GCN}, H_{emb})$$

where $F \in R^{l \times (4*dim)}$ denotes the multi-level features, $O \in R$ and $O \in R$ refer to the output of the Bi-LSTM and GCN respectively and $H \in R$ denotes the embedding matrix of Bert.

Convolutional neural network layer. Convolutional neural network (CNN) utilizes several kernels or filters to extract local features of the given matrix. In a single convolution operation, a filter, or a kernel is computed as follows:

$$x_i = f(k \cdot f_{i:i+j-1} + b)$$

where $k \in R^{s \times (4*dim)}$ denotes the 1D convolutional kernel we adopted, b denotes the bias of the convolutional kernel and f denotes the concatenation of $[f, f, \ldots\ldots, f_{i+j}]$ that means a single convolution operation involves several continuous feature vectors. A single convolution slides in a given feature matrix, and finally obtains a vector of deeper features.

Classification layer. In this section, we show how to classify the sentiment based on the vector obtained by the convolution operation. We apply a linear layer followed by a *SoftMax* layer to classify the vector. It can be described as follows:

$$L = softmax(\bar{X}W_L + b_L)$$

where $\bar{X}$ refers to the vector obtained in Part 4, W and b refers to the weights and bias of the linear layer, respectively, and L refers to the predicted results of the given labels.

Experiment and results

Experiment settings

In our experiments, we used some baseline to compare the generated performances. Three classic machine learning models were used,[39] logistic regression (LR), support vector machines (SVMs) and random forests (RFs). In addition, there are three ensemble learning,[40] namely adaboost (AB), bagging (BG) and XGBoost (XGB). Meanwhile, we used seven deep learning models; the specific models are as follows:

TextCNN[41]: Uses a trainable embedding to gain word vectors and classify the vaccine stance by 1 dimensional filter.

BERT: Performs fine-tuning based on the pre-trained model and then models it into a classifier.

BERT-CNN[42]: Initializes word embeddings by Bert and classifies the given sentences by dimensional filters.

BERT-GCN-CNN (BertGCTM)[41,43]: Initializes the word embeddings through the Bert model and then inputs the matrix to the Graph Neural Network followed by CNN. Finally, it classifies the feature matrix to the three vaccine stances.

BERT-GCN-LSTM (BertGCTM)[43,44]: Initializes the word embeddings through the Bert model. After that, it feeds the feature matrix into the Graph Neural Network and then models the derived output of GNN to LSTM. Finally, it classifies the hidden states generated by LSTM.

BERT-LSTM-attention (BertATTM)[45]: Initializes the word embeddings through the Bert model. Next, it models the obtained embeddings directly to LSTM. After that, it leverages multi-head attention layer and linear layer to gain the classification consequences.

BERT-multi-channel-CNN (BertMCNN)[42,46]: Initializes the word embeddings through the Bert model. Then, it leverages the proposed multi-channel feature extraction method to obtain different kinds of features based on the Graph Neural Network and the Bi-LSTM. Next, it concatenates the embeddings, the features derived from GNN and the features from the Bi-LSTM and feeds the concatenation to the CNN followed by a classifier.

Note that, for the machine learning algorithm, we used the default parameters from the scikit-learn package in Python. For the deep learning algorithm, we utilized the Bert tokenizer to tokenize the word for all models mentioned above and we adopted the base type of the Bert model, which means the dimension of hidden states and word embeddings is 768. For the CNN component, 10 different sizes of kernels $[1, 2, 3 \ldots, 10]$ were used to extract the information. In addition, we set the learning rate as 5×10^{-5}, batch size as $\{4, 8, 16\}$, dropout rate as 0.1 and max train epochs as 40. We adopted accuracy as the evaluation metric to test the performances of all baseline, ablation, and proposed models. All our deep learning models were implemented using the PyTorch framework.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Results

Classification results. To compare the performance of all models, we aggregated the results in Tables 4 and 5. It is evident that deep learning algorithms significantly outperformed traditional machine learning methods. Among them, our proposed model achieved the best or notably good results across all four datasets. Specifically, in the

Table 4. The results of using different algorithms by four labels for different brands.

Method	Pfizer (%)				AstraZeneca (%)				Sinovac (%)			
	Att	Will	Text + att	Text + will	Att	Will	Text + att	Text + will	Att	Will	Text + att	Text + will
LR	59.69	76.44	76.44	60.21	51.45	79.16	78.89	52.77	46.02	68.47	67.33	52.56
SVM	59.95	76.96	77.23	62.83	49.60	79.42	79.68	50.66	47.73	66.19	65.63	51.70
RF	57.33	77.49	76.96	59.69	50.13	78.36	78.63	54.62	45.74	66.48	65.34	53.41
AB	53.93	71.47	74.35	53.40	43.80	70.71	72.03	44.59	45.17	55.68	59.66	44.60
BG	57.33	79.84	78.80	56.81	51.19	79.16	79.16	54.62	48.86	64.49	62.50	52.84
XGB	57.33	74.61	76.44	57.85	46.70	76.78	78.63	50.40	43.18	63.92	66.48	48.86
TextCNN	80.84	93.44	94.75	80.32	80.69	89.42	91.27	80.95	76.64	84.62	86.33	83.48
Bert	84.51	94.75	93.70	82.68	84.92	92.59	94.44	83.07	81.20	84.05	86.90	83.76
BertCNN	84.25	93.96	93.96	82.68	83.33	93.11	93.92	81.75	82.91	83.42	84.33	82.05
BertGCNN	84.74	93.95	92.90	83.95	83.60	94.18	95.50	83.60	82.76	82.47	84.90	84.05
BertGCTM	83.16	93.68	93.42	83.42	80.95	94.18	94.71	81.75	81.90	82.18	82.76	84.62
BertATTM	85.00	94.21	94.47	82.68	84.39	94.44	93.92	82.28	78.45	83.33	84.48	83.33
BertMCNN	84.51	95.28	95.01	85.04	84.39	94.71	95.24	83.33	82.05	84.05	86.61	84.62

context of the twelve classification tasks for the three vaccines, our BertMCNN model delivered the top results in five tasks while securing second or third place in the remaining seven tasks. Obviously, the Bert-based model yields higher result scores, indicating a substantial performance gap between the Bert-based model and other models. Specifically, when comparing TextCNN and BertCNN, it becomes apparent that BertCNN outperforms the TextCNN model significantly. This highlights the efficiency of Bert in converting words into vectors, which proves crucial in text classification tasks.

Table 5 presents the F1-scores of all models, where, like the accuracy shown above, the deep learning model demonstrates its outstanding performance compared to the machine learning algorithm. What is worth noting is our proposed model's consistent superiority. Across the three datasets, encompassing a collective tally of 12 distinct classification tasks, our proposed model manifests its efficiency by securing the highest scores while attaining second or third position in the remaining six tasks. These outcomes serve as robust evidence affirming the encouraging performance of our devised model.

Moreover, the results produced by BertCNN are only marginally inferior to those of Bert on three datasets. This

discrepancy might arise because CNN can only capture local features within a limited range, whereas textual features often exhibit a global nature. CNNs might need help with handling sequence vectors effectively. When comparing the proposed model with the BertGCNN model, it becomes evident that the BertGCNN model performs admirably, achieving results only marginally lower than the proposed model on individual tasks and even slightly surpassing the proposed BertMCNN model on some tasks. This result indicates that GCN and CNN layers are used together to model the text sequence's contextualized information. Overall, the GCN models exhibited in the table outperform the other models. This result underscores the significance of GCN's ability to construct the graph from the dependency tree, enabling it to fully capture semantic information and enhance performance in subsequent components. In the overall context, our proposed deep learning model incorporating causality plays a vital role. When considering the prediction accuracy for the three vaccine brands, the overall accuracy achieved through reasoning inputs is notably higher than the mere 3% achieved with single-label inputs. This improvement underscores the significant contribution of incorporating causality to enhance the results.

Table 5. The F1-score of using different algorithms by four labels for different brands.

Method	Pfizer (%)				AstraZeneca (%)				Sinovac (%)			
	Att	Will	Text + att	Text + will	Att	Will	Text + att	Text + will	Att	Will	Text + att	Text + will
LR	44.40	44.40	44.40	44.40	45.96	35.95	35.01	47.66	37.92	42.26	45.41	45.95
SVM	45.26	45.26	45.26	45.26	45.33	37.51	37.61	47.10	41.48	40.10	45.18	45.99
RF	44.40	44.40	44.40	44.40	42.05	33.16	33.28	48.77	35.86	38.84	38.09	49.71
AB	45.26	45.26	45.26	45.26	41.65	40.76	40.66	42.89	43.36	38.63	48.24	43.42
BG	44.40	44.40	44.40	44.40	46.88	39.66	39.19	51.24	43.63	39.71	49.36	49.81
XGB	45.26	45.26	45.26	45.26	45.47	37.13	37.92	48.41	39.16	42.99	52.71	52.23
TextCNN	76.91	91.36	92.39	76.54	79.90	73.91	81.50	80.26	75.80	78.70	84.50	83.48
Bert	82.05	92.28	86.18	79.77	84.46	82.57	87.81	82.73	81.42	78.71	83.88	83.93
BertCNN	82.02	86.40	91.57	80.39	83.26	83.11	85.83	81.62	82.96	78.61	79.58	82.40
BertGCNN	82.57	88.71	87.80	81.80	83.52	88.85	87.67	83.47	82.70	78.09	83.25	84.67
BertGCTM	80.21	91.35	91.10	81.38	80.33	86.97	88.06	81.07	81.67	76.53	80.11	84.84
BertATTM	82.85	91.87	92.07	80.47	84.15	87.10	84.34	81.85	77.91	78.93	82.01	83.47
BertMCNN	82.38	92.71	92.44	83.26	84.14	88.45	88.92	83.24	82.24	78.15	84.38	84.71

Further analysis

In addition to analyzing the model's performance, as mentioned earlier, we can also glean valuable insights from the prediction results of different vaccines. Overall, while there is some variation in outcome predictions among other algorithms, they consistently show lower accuracy for Sinovac and higher accuracy for AstraZeneca and Pfizer vaccines, as shown in Figure 4. This result reflects the public's views and attitudes towards each brand in different regions to a certain extent.

Our data are for the whole year of 2022. This period can be referred to as the post-epidemic era, during which people's attitudes towards virus control have transformed compared to the early stages of the outbreak. People now strive to balance between epidemic preparedness and returning to everyday life. This year, China continued to implement strict measures to prevent and control epidemics, and despite a significant portion of the population being vaccinated following expert recommendations, there were still instances of large-scale epidemic spread. These occurrences have led to public skepticism regarding the effectiveness of the Sinovac vaccine, consequently fostering a relatively negative perception of it. On the other hand, in the United Kingdom and the United States, where Pfizer and AstraZeneca vaccines are more widely administered, the spread of epidemics occurred in recurring waves. As a result, the public has become more accustomed to this situation and does not have excessive expectations of vaccines. Consequently, their attitude towards vaccines remains relatively stable, and they do not experience significant fluctuations in response to specific rounds of epidemics.

To further explore the predictive ability of our proposed model, we analyzed its performance concerning two distinct tasks: the prediction of Attitude and Willingness from textual inputs without the other label. Figure 5 shows the confusion matrix of the predictive outcomes for these two tasks. For the Attitude prediction task, it is evident that there is no discernible bias within the proposed model, while for the Willingness prediction task, the model tends to predict the user's willingness toward 'Not Sure'. This tendency comes from the notable prevalence of 'Not Sure' instances within the dataset. Furthermore, although we have oversampled the text under the 'No' label, there is still an upside in the model's prediction for this text category.

We selected a snippet of text data, "More rare side effects from AstraZeneca vaccine." from our dataset and fed it into the BERT model discussed in Figure 6. Within the BERT model lies an attention mechanism, the weighted
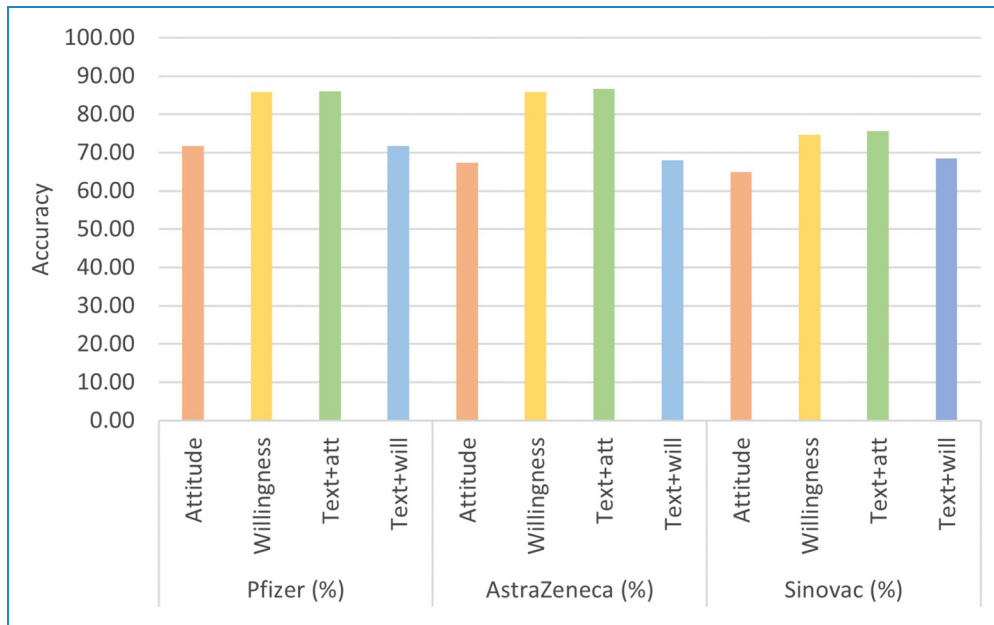
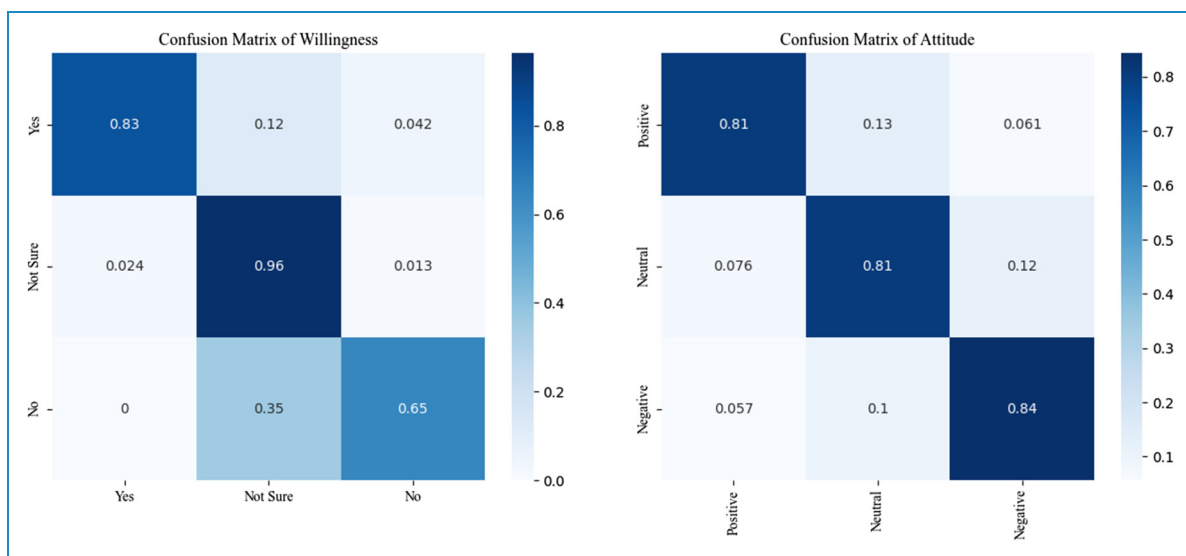**Figure 4.** The average accuracy performance of each brand.

Figure 5. The confusion matrix of the prediction of attitude and willingness from text input without the other label.

results we have visualized. The visualization illustrates the degree of correlation between pairs of tokens; the more robust their correlation, the higher the weight value. For instance, consider the words "side" and "effects" depicted in the figure. Their conjunction forms the term "side effects," a pivotal phrase in vaccine evaluation, warranting a correspondingly high weight, as emphasized in the visualization.

Conversely, when traditional statistical machine learning approaches are employed, methods like TF-IDF, which rely on statistical feature extraction, are typically utilized. Such techniques can account for the frequency of specific word elements

but must catch up in capturing the nuanced positional and semantic relationships between them. This nuanced understanding of context and meaning is where deep learning models excel, showcasing their advantage over traditional methodologies.

Discussions

Implications for research

Understanding vaccine hesitancy is crucial for effective public health interventions and disease surveillance.[47]
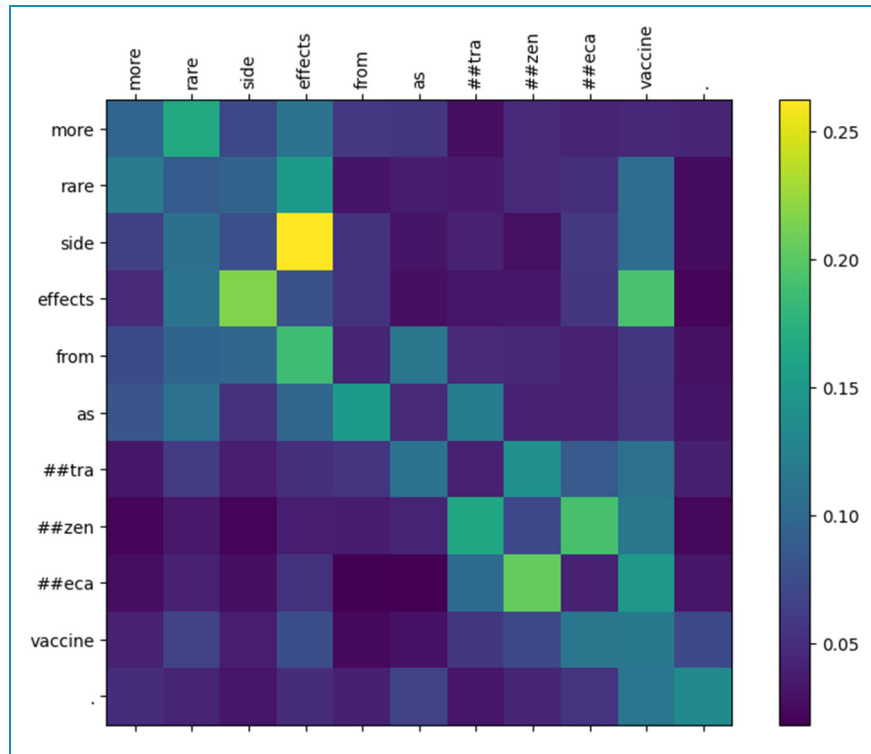
Figure 6. Text weight heatmap.

Most existing studies analyzing vaccine hesitancy on social media focus on global perspectives (positive, negative, neutral) or anti-vaccine issues. In this study, people's underlying attitudes towards vaccines in post epidemic era are considered rather than categorized as solely anti- or pro-vaccine.[48] This approach acknowledges the heterogeneous nature of vaccine-hesitant individuals along a continuum, making it challenging to identify and label such content accurately. Our study distinguishes itself from previous research by offering a complementary approach that enhances recognition accuracy and develops strategies to mitigate vaccination hesitancy.[49] By considering the spectrum of vaccine hesitancy, our causality deep learning techniques can capture a more nuanced understanding of people's attitudes towards vaccination. This enables the creation of a robust and adaptable framework for identifying vaccine hesitancy in social media content, which is particularly relevant in the dynamic landscape of the post-epidemic era.

The implications of our research extend beyond academia and can be valuable in government administration for disease surveillance and outbreak communication.[2] By effectively identifying vaccine hesitancy, health authorities can tailor communication strategies to address the concerns of hesitant individuals and encourage vaccination. Moreover, our approach allows for the integration of digital technologies in outbreak management and production systems,[50] to minimize the impact of outbreaks on

society and economy. By harnessing the power of deep learning, we can respond swiftly to emerging health challenges and mitigate their consequences.

Limitations and future work

This study has identified several limitations that must be addressed in future research. Firstly, the tweets used in this paper constitute only a tiny fraction of daily tweets due to Twitter's data access restrictions. Consequently, the messages collected may only partially represent a global perspective.[51] To overcome this limitation, future work should incorporate tweets from relevant regions to ensure a more comprehensive and diverse dataset. Another limitation pertains to the fine-tuning process of the models, which were restricted to specific hyperparameters such as learning rate, batch size, and number of periods. This limitation led to the neglect of other important parameters that could have enhanced the models' performance. Thus, further research should focus on extensive tuning of the models to optimize their effectiveness in identifying vaccine-hesitant tweets for future studies.

As for future directions, additional investigations should be conducted to explore the role of social media in exacerbating hesitancy about the COVID-19 vaccine. To delve deeper into this aspect, employing content analysis and network analysis techniques such as topic modeling, trend analysis, and social network analysis.[52] These methods can provide valuable

insights into the dynamics of hesitancy-related discussions on social media. Furthermore, understanding the extent to which the rhetoric of vaccine hesitancy influences vaccination attitudes and behaviors within the population is crucial.[53] This can be achieved by studying how vaccine hesitancy information spreads among communities of social media users.

Conclusion

Vaccine hesitancy has always existed in the post-epidemic era, but not global anti-vaccine sentiment, its causes, and consequences. This study aims to identify hesitant behaviors in public vaccination through deep learning models. This paper presents an innovative methodology focusing on social media monitoring and deep learning models to address vaccine hesitancy effectively. By employing these advanced algorithms, public health authorities and organizations can proactively combat the adverse impact of vaccine hesitancy on vaccine acceptance rates. This approach offers valuable insights and lessons in assessing public opinion, thereby enhancing decision-making processes related to vaccine deployment. As we navigate the complexities of the post-epidemic era, where attitudes towards virus control and vaccination have shifted, our research paves the way for more informed and data-driven initiatives to promote vaccine acceptance in the post-epidemic era.

ORCID iD: Yang Liu https://orcid.org/0000-0002-9410-1755

References

1. Dong L, Hu S and Gao J. Discovering drugs to treat coronavirus disease 2019 (COVID-19). *DD&T* 2020; 14: 58–60.
2. Liu Y, Shi J, Zhao C, et al. Generalizing factors of COVID-19 vaccine attitudes in different regions: a summary generation and topic modeling approach. *DIGITAL HEALTH* 2023; 9: 20552076231188852.
3. del Rio C, Omer SB and Malani PN. Winter of omicron—the evolving COVID-19 pandemic. *JAMA* 2022; 327: 319.
4. Troiano G and Nardi A. Vaccine hesitancy in the era of COVID-19. *Public Health* 2021; 194: 245–251.
5. Loomba S, de Figueiredo A, Piatek SJ, et al. Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nat Hum Behav* 2021; 5: 337–348.
6. Dubé E, Laberge C, Guay M, et al. Vaccine hesitancy: an overview. *Hum Vaccin Immunother* 2013; 9: 1763–1773.
7. Linas BP, Savinkina A, Barbosa C, et al. A clash of epidemics: impact of the COVID-19 pandemic response on opioid overdose. *J Subst Abuse Treat* 2021; 120: 108158.
8. Filho CIS. Using mobile phone-based text message to recruit representative samples: assessment of a cross-sectional survey about the COVID-19 vaccine hesitation. *Int J Med Inf* 2022; 165: 104832.
9. Roozenbeek J, Schneider CR, Dryhurst S, et al. Susceptibility to misinformation about COVID-19 around the world. *R Soc Open sci* 2020; 7: 201199.
10. Wawrzuta D, Jaworski M, Gotlib J, et al. What arguments against COVID-19 vaccines run on Facebook in Poland: content analysis of comments. *Vaccines* 2021; 9: 481.
11. Tsao S-F, Chen H, Tisseverasinghe T, et al. What social media told us in the time of COVID-19: a scoping review. *Lancet Digital Health* 2021; 3: e175–e194.
12. Liu Y, Chi M and Sun Q. Sarcasm detection in hotel reviews: a multimodal deep learning approach. *JHTT* 2024. DOI: 10.1108/JHTT-04-2023-0098.
13. Liu Y, Shi J, Huang F, et al. Unveiling consumer preferences in automotive reviews through aspect-based opinion generation. *J Retailing Consum Serv* 2024; 77: 103605.
14. Guo Y, Zhang Y, Lyu T, et al. The application of artificial intelligence and data integration in COVID-19 studies: a scoping review. *J Am Med Inform Assoc* 2021; 28: 2050–2067.
15. Qorib M, Oladunni T, Denis M, et al. Covid-19 vaccine hesitancy: text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset. *Expert Syst Appl* 2023; 212: 118715.
16. Muric G, Wu Y and Ferrara E. COVID-19 vaccine hesitancy on social media: building a public Twitter data set of antivaccine content, vaccine misinformation, and conspiracies. *JMIR Public Health Surveill* 2021; 7: e30642.
17. Liu Y, Fei H, Zeng Q, et al. Electronic word-of-mouth effects on studio performance leveraging attention-based model. *Neural Comput Appl* 2020; 32: 17601–17622.
18. Yao L, Chu Z, Li S, et al. A survey on causal inference. *ACM Trans Knowl Discov Data* 2021; 15: 1–46.
19. Li X, Zhou M, Wu J, et al. Analyzing COVID-19 on Online Social Media: Trends, Sentiments and Emotions, http://arxiv.org/abs/2005.14464 (2020, accessed 10 December 2023).

20. Zhou J, Zogan H, Yang S, et al. Detecting community depression dynamics due to COVID-19 pandemic in Australia. *IEEE Trans Comput Soc Syst* 2021; 8: 982–991.

21. Yin H, Yang S, Li J, et al. Detecting topic and sentiment dynamics due to COVID-19 pandemic using social Media. In: Yang X, Wang C-D and MdS I (eds) *Advanced data mining and applications*. Cham: Springer International Publishing, pp. 610–623.

22. Department of Computer Science Engineering, BBD University, Lucknow, India, Mishra A. A comprehensive analysis of approaches for sentiment analysis using Twitter data on COVID-19 vaccines. *J Infor Electr Electron Eng* 2021; 2: 1–10.

23. Lyu JC, Han EL and Luli GK. COVID-19 vaccine–related discussion on Twitter: topic modeling and sentiment analysis. *J Med Internet Res* 2021; 23: e24435.

24. Liu S and Liu J. Public attitudes toward COVID-19 vaccines on English-language Twitter: a sentiment analysis. *Vaccine* 2021; 39: 5499–5505.

25. Mehedi Shamrat FMJ, Chakraborty S, Imran MM, et al. Sentiment analysis on twitter tweets about COVID-19 vaccines usi ng NLP and supervised KNN classification algorithm. *IJEECS* 2021; 23: 463.

26. Khan R, Rustam F, Kanwal K, et al. US Based COVID-19 tweets sentiment analysis using TextBlob and supervised machine learning algorithms. In: 2021 International Conference on Artificial Intelligence (ICAI), Islamabad, Pakistan, pp.1–8: IEEE.

27. Raza GM, Butt ZS, Latif S, et al. Sentiment analysis on COVID tweets: an experimental analysis on the impact of count vectorizer and TF-IDF on sentiment predictions using deep learning models. In: 2021 International Conference on Digital Futures and Transformative Technologies (ICoDT2), Islamabad, Pakistan, pp.1–6: IEEE.

28. Wang G and Kwok SWH. Using K-means clustering method with Doc2Vec to understand the twitter Users' opinions on COVID-19 vaccination. In: 2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI), Athens, Greece, pp.1–4: IEEE.

29. Bonnevie E, Gallegos-Jeffrey A, Goldbarg J, et al. Quantifying the rise of vaccine opposition on Twitter during the COVID-19 pandemic. *J Commun Healthc* 2021; 14: 12–19.

30. Tavoschi L, Quattrone F, D'Andrea E, et al. Twitter as a sentinel tool to monitor public opinion on vaccination: an opinion mining analysis from September 2016 to August 2017 in Italy. *Hum Vaccin Immunother* 2020; 16: 1062–1069.

31. Hussain A, Tahir A, Hussain Z, et al. Artificial intelligence–enabled analysis of public attitudes on Facebook and Twitter toward COVID-19 vaccines in the United Kingdom and the United States: observational study. *J Med Internet Res* 2021; 23(4): e26627.

32. Devaraj J, Madurai Elavarasan R, Pugazhendhi R, et al. Forecasting of COVID-19 cases using deep learning models: is it reliable and practically significant? *Results Phys* 2021; 21: 103817.

33. Shastri S, Singh K, Kumar S, et al. Time series forecasting of Covid-19 using deep learning models: India-USA comparative case study. *Chaos, Solitons Fractals* 2020; 140: 110227.

34. Poria S, Cambria E, Bajpai R, et al. A review of affective computing: from unimodal analysis to multimodal fusion. *Inf Fusion* 2017; 37: 98–125.

35. Devlin J, Chang M-W, Lee K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv:181004805 [cs]*, http://arxiv.org/abs/1810.04805 (2019, accessed 16 April 2022).

36. Manning C, Surdeanu M, Bauer J, et al. The Stanford CoreNLP natural language processing toolkit. In: Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, Baltimore, Maryland, pp.55–60: Association for Computational Linguistics.

37. Kipf TN and Welling M. Semi-Supervised Classification with Graph Convolutional Networks, http://arxiv.org/abs/1609.02907 (2017, accessed 26 July 2023).

38. Liu Y, Ding X, Peng S, et al. Leveraging ChatGPT to optimize depression intervention through explainable deep learning. *Front Psychiatry* 2024; 15: 1383648.

39. Jordan MI and Mitchell TM. Machine learning: trends, perspectives, and prospects. *Science* 2015; 349: 255–260.

40. Sagi O and Rokach L. Ensemble learning: a survey. *WIREs Data Mining Knowl Discov* July 2018; 8(4): e1249. DOI: 10.1002/widm.1249.

41. Kim Y. Convolutional Neural Networks for Sentence Classification. *arXiv:14085882 [cs]*, http://arxiv.org/abs/1408.5882 (2014, accessed 16 April 2022).

42. Dong J, He F, Guo Y, et al. A commodity review sentiment analysis based on BERT-CNN model. In: 2020 5th International Conference on Computer and Communication Systems (ICCCS), Shanghai, China, pp.143–147: IEEE.

43. Lin Y, Meng Y, Sun X, et al. BertGCN: Transductive Text Classification by Combining GCN and BERT, http://arxiv.org/abs/2105.05727 (2022, accessed 26 July 2023).

44. Graves A, Jaitly N and Mohamed A. Hybrid speech recognition with deep bidirectional LSTM. In: 2013 IEEE Workshop on Automatic Speech Recognition and Understanding, Olomouc, Czech Republic, pp.273–278: IEEE.

45. Ezen-Can A. A Comparison of LSTM and BERT for Small Corpus, http://arxiv.org/abs/2009.05451 (2020, accessed 26 July 2023).

46. Sun W, Min X, Zhai G, et al. MC360IQA: a multi-channel CNN for blind 360-degree image quality assessment. *IEEE J Sel Top Signal Process* 2020; 14: 64–77.

47. Liu Y. Depression clinical detection model based on social media: a federated deep learning approach. *J Supercomput* 2024; 80(6): 7931–7954. DOI: 10.1007/s11227-023-05754-7.

48. Dubé È, Ward JK, Verger P, et al. Vaccine hesitancy, acceptance, and anti-vaccination: trends and future prospects for public health. *Annu Rev Public Health* 2021; 42: 175–191.

49. Alamoodi AH, Zaidan BB, Al-Masawa M, et al. Multi-perspectives systematic review on the applications of sentiment analysis for vaccine hesitancy. *Comput Biol Med* 2021; 139: 104957.

50. Gupta M, Shoja A and Mikalef P. Toward the understanding of national culture in the success of non-pharmaceutical technological interventions in mitigating COVID-19 pandemic. *Ann Oper Res* 2022; 319: 1433–1450.

51. Liu Y, Ding X, Chi M, et al. Assessing the helpfulness of hotel reviews for information overload: a multi-view spatial feature approach. *Inf Technol Tourism* 2024; 26: 59–87.

52. Karami A, Zhu M, Goldschmidt B, et al. COVID-19 vaccine and social media in the U.S.: exploring emotions and discussions on Twitter. *Vaccines* 2021; 9: 1059.

53. Liu Y. Analyzing the effect of user-generated content on studio performance: a combined approach. *Manage Decis Econ* 2024; 45: 2228–2248.