Review

Role of Artificial Intelligence for Analysis of COVID-19 Vaccination-Related Tweets: Opportunities, Challenges, and Future Trends

Wajdi Aljedaani [1] , **Eysha Saad** [2] , **Furqan Rustam** [3] , **Isabel de la Torre Díez** [4,*] **and Imran Ashraf** [5,*]

[1] Department of Computer Science and Engineering, University of North Texas, Denton, TX 76203, USA
[2] Faculty of Computer Science and Information Technology, Khawaja Fareed University of Engineering and Information Technology, Rahim Yar Khan 64200, Pakistan
[3] Department of Software Engineering, School of Systems and Technology, University of Management and Technology Lahore, Lahore 54770, Pakistan
[4] Department of Signal Theory and Communications and Telematic Engineering, Unviersity of Valladolid, Paseo de Belén 15, 47011 Valladolid, Spain
[5] Department of Information and Communication Engineering, Yeungnam University, Gyeongsan 38541, Korea
[*] Correspondence: isator@tel.uva.es (I.d.l.T.D.); imranashraf@ynu.ac.kr (I.A.)

Abstract: Pandemics and infectious diseases are overcome by vaccination, which serves as a preventative measure. Nevertheless, vaccines also raise public concerns; public apprehension and doubts challenge the acceptance of new vaccines. COVID-19 vaccines received a similarly hostile reaction from the public. In addition, misinformation from social media, contradictory comments from medical experts, and reports of worse reactions led to negative COVID-19 vaccine perceptions. Many researchers analyzed people's varying sentiments regarding the COVID-19 vaccine using artificial intelligence (AI) approaches. This study is the first attempt to review the role of AI approaches in COVID-19 vaccination-related sentiment analysis. For this purpose, insights from publications are gathered that analyze the (a) approaches used to develop sentiment analysis tools, (b) major sources of data, (c) available data sources, and (d) the public perception of COVID-19 vaccine. Analysis suggests that public perception-related COVID-19 tweets are predominantly analyzed using TextBlob. Moreover, to a large extent, researchers have employed the Latent Dirichlet Allocation model for topic modeling of Twitter data. Another pertinent discovery made in our study is the variation in people's sentiments regarding the COVID-19 vaccine across different regions. We anticipate that our systematic review will serve as an all-in-one source for the research community in determining the right technique and data source for their requirements. Our findings also provide insight into the research community to assist them in their future work in the current domain.

1. Introduction

The outbreak of the COVID-19 disease caused by the novel severe acute respiratory syndrome coronavirus 2 (SARS-COV-2) in December 2019 has strongly affected the mental health and life of people around the world. As of 19 August 2022, COVID-19 has infected more than 591 million people on a global scale with more than 6 million deaths [1]. The alarming increase in the number of infected cases, as well as the numerous genetic variations of coronavirus, sparked a major global research and development (R&D) effort to develop a vaccine to mitigate the threat of this infectious disease. Governments, scientists, and pharmaceutical companies all over the world were determined to develop effective vaccines.

Multiple candidates of COVID-19 vaccines had been developed over a time period of 6–8 months, beginning on 8 April 2020 [2]. A few candidates, such as the ChAdOx1 vaccine developed by AstraZeneca/Oxford and messenger RNA (mRNA) vaccines by Pfizer/BioNTech and Moderna, have been authorized globally [3]. Despite the exceptional results of vaccines in clinical trials in terms of efficacy and safety, "vaccine hesitancy" [4] towards authorized candidates of COVID-19 remains a significant challenge in achieving herd immunity [5].

Vaccine hesitancy has been reported as the most substantial threat to the health industry as it reduces the vaccine acceptance rate, thus resulting in the recurrence of pandemics of infectious diseases [6]. For instance, in the US and Europe, measles outbreaks escalated rapidly due to the refusal of individuals to get vaccinated against the infection [7,8]. Unwillingness to uptake vaccine against poliovirus [9–11] or human papillomavirus [12] have also been recorded in various countries. A recent systematic review of 63 surveys found that around 66.1% of the global population showed a willingness to take the COVID-19 vaccine [13]. Consequently, another review study confirmed the significant influence of social media platforms on vaccine hesitancy [14].

Social media platforms have increasingly become a significant source for analyzing the public's perception of an entity as they allow people from around the world to participate in public discussions. These online networking platforms have been actively regarded as important tools for real-time surveillance of infectious illnesses and vaccine acceptance or hesitancy [15]. Planning effective health communication to encourage vaccine uptake and successfully implementing herd immunization requires a thorough understanding of the public's perception and potential determinants of people's opinions towards the COVID-19 vaccine. Analysis of underlying sentiments of users' opinions, attitudes, and comments by integrating sentiment analysis techniques seems promising in categorizing the public's perception concerning the COVID-19 vaccine [16].

Sentiment analysis is the automated extraction and analysis of subjective opinions on a variety of aspects of an entity. It integrates Natural Language Processing (NLP) techniques with the aim of identifying the emotions of the text writer via the written words [17]. It computationally differentiates and analyzes the writer's opinion about the topic upon which assumption is based in a specific segment of text. The fundamental goal of sentiment analysis is to determine the degree of polarity and to classify the sentiment communicated by the writer in a text as positive, negative, or neutral.

1.1. Existing Reviews on COVID-19

Several studies have been conducted on COVID-19 for systematic literature review and survey [18]. For example, Shah et al. [19] performed a comprehensive review on COVID-19 detection approaches using medical images. The study focused on X-ray and CT lung images. Lalmuanawma et al. [20] reviewed machine learning and artificial intelligence applications used to predict, forecast, and develop medicine for SARS-CoV-2. The study [21] covers the computer vision-based approach to control COVID-19. Similarly, studies reviewing the use of cloud computing in education during the COVID-19 pandemic [22], the analytical performance of COVID-19 detection methods [23], and prediction models with respect to age, gender, and pre-existing medical conditions [24] can be found. A similar study to ours reviews the literature on high-income countries and COVID-19 vaccine hesitancy [25]; however, its scope is short as it covers a limited number of countries. Contrarily, the initial pool of publications in the current study is not restricted to any particular geographic area. Another distinction lies in the fact that this study reviews the articles that perform sentiment analysis regarding people's perceptions of COVID-19 vaccination. Alamoodi et al. [26] performed reviews on sentiment analysis in fighting COVID-19 and infectious diseases. However, none of the previous studies review sentiment analysis using machine learning and artificial intelligence for understanding and predicting the public's perception of COVID-19 vaccination. To the best of our knowledge, this is the first study to

review and analyze articles about the public perception of COVID-19 vaccination using machine learning and artificial intelligence applications.

1.2. Scope of Study

Although several studies have proposed solutions to screen people's sentiments towards the COVID-19 vaccine, to the best of our knowledge, no comprehensive reviews have been undertaken to synthesize and uncover the role of artificial learning (AI) and machine learning (ML) in screening the public's sentiment towards COVID-19 vaccines. Therefore, this study adopts a systematic mapping approach to extensively review the progress of AI and ML techniques employed to analyze and investigate people's sentiments towards the COVID-19 vaccine. This study aims to systematically summarize the workflow of existing studies, sum up the frequently used methods for the automated screening of peoples' sentiments, and collect all the multiple sources of datasets of public reviews on the COVID-19 vaccine so that novice researchers can determine a correct tool for a better solution.

1.3. Research Questions

To offer a deeper insight into the applications of AI and ML techniques in the screening of people's perceptions of the COVID-19 vaccine and to achieve our research goal, we propose the following research questions (RQs):

1.3.1. RQ_1: Which Approaches Have Been Used in the Development of Sentiment Analysis Tools for Screening the Public Perception of COVID-19 Vaccines?

With this question, we want to provide insights into the existing AI and ML practices employed in the screening process of public sentiments towards the COVID-19 vaccine. In this way, we obtain an overview of the existing approaches, which will help the research community in determining the best solution for future research. RQ_1 is further subdivided into two more RQs:
RQ_1-a: Which sentiment classifiers have been frequently used?
RQ_1-b: Which feature engineering technique has the dominant role for sentiment analysis?

1.3.2. RQ_2: What Are the Major Sources of Data to Monitor People's Opinions towards COVID-19 Vaccines?

We aim at investigating the data utilized to train and evaluate the AI and ML-based sentiment analysis tools. Additionally, we gain an overview of datasets that are publicly available to assist the research community for future research. Hence, proposing sub-RQ_2s:
RQ_2-a: What is the availability status of datasets utilized to assess the public's sentiments regarding COVID-19 vaccines?
RQ_2-b: Which techniques are employed to annotate the unlabeled data records according to their sentiments?

1.3.3. RQ_3: What Is the Public Perception of the COVID-19 Vaccine According to the Reviewed Studies' Results? What Are the Geographical Locations Where Public Perception Has Been Studied Regarding COVID-19 Vaccines?

With this question, we gain an overview of people's opinions or feelings regarding the COVID-19 vaccine. Another pertinent point in this question is the discussion of geographical locations where public perception has been studied regarding COVID-19 vaccines.

1.4. Contributions

With this systematic review, we provide the research community with information regarding the role of AI and ML in monitoring the public perspective of the COVID-19 vaccine over the span of 2 years. Our main contributions are outlined below:

- A catalog of 47 peer-reviewed publications that employ AI and ML techniques to analyze the public perceptions of COVID-19 vaccines is systematically reviewed;

- We highlight the approaches used for the construction of sentiment analysis tools regarding the sentiment analysis of COVID-19-related text;
- Insights regarding the major data sources and their availability status are provided;
- Public perception regarding the COVID-19 vaccine is discussed according to the selected pool of publications;
- The underlying geographical locations of studies are highlighted;
- A discussion of how our findings can be used to improve future research in this area and the factors to consider when choosing an approach to monitor public perception regarding the COVID-19 vaccine.

The remainder of this paper is structured into five sections. The systematic approach followed for data collection, sorting, and inclusion and exclusion of irrelevant material are presented in Section 2. The results are discussed and interpreted with regard to each formulated research question in Section 3. Section 4 presents the discussions of findings, followed by recommendations in Section 5. In the end, the study is concluded in Section 6.

2. Materials and Methods

Our research employs a systematic mapping approach to comprehensively review the literature related to the public's perception of the COVID-19 vaccine using AI techniques. We followed the guidelines of a systematic mapping study (SMS), which is a means of unbiased cataloging and summarizing relevant information regarding our *RQ*s that provides readers with deeper understanding and insights into the underlying topic [27–30]. To achieve this goal, we followed a three-step strategy:

1. Planning: Publications are identified, screened, and validated from digital libraries based on inclusion/exclusion criteria.
2. Execution: Publications are read to filter out irrelevant studies.
3. Synthesis: Classification and analysis of extracted data to answer the designed *RQ*s.

2.1. Planning

The literature search in this study covers publications from peer-reviewed journals that are indexed in seven well-known digital libraries (listed in Table 1). These digital libraries are preferred for this review based on their scientific soundness and reliability, and they are considered sufficient and adequate.

Table 1. The digital libraries used for finding articles in this study.

No.	Digital Library	URL
1	ACM Digital Library	https://dl.acm.org/, accessed on 15 December 2021
2	IEEE Xplore	https://ieeexplore.ieee.org/, accessed on 15 December 2021
3	Science Direct	https://www.sciencedirect.com/, accessed on 15 December 2021
4	scopus	https://www.scopus.com/, accessed on 15 December 2021
5	PubMed	https://pubmed.ncbi.nlm.nih.gov/, accessed on 15 December 2021
6	MedRxiv	https://www.medrxiv.org/, accessed on 15 December 2021
7	Web of Science	https://webofknowledge.com/, accessed on 15 December 2021

2.1.1. Inclusion and Exclusion Criteria

Inclusion and exclusion criteria are significant to reduce bias, prune the search area, retrieve relevant publications, and eliminate studies that do not contribute to the *RQ*s. The filtered peer-reviewed publications that meet these criteria help reviewers in manual filtering to see whether these publications are appropriate for the research, i.e., adopted or proposed AI techniques for sentiment analysis of public perception regarding the COVID-19 vaccine. We also utilized the initial pool of publications for backward and forward snowballing. The utilized inclusion and exclusion criteria in this study are summarized in Table 2. Regarding the time period, the starting date was set to 1 January 2020, and the end

date was set to 31 December 2021, which allows the selection of publications that appeared before 1 January 2022.

Table 2. Inclusion and exclusion criteria followed in this study.

Inclusion Criteria
1. Published in 2020 or 2021;   2. Written in English;   3. Available in digital format;   4. Proposed or used AI or ML or deep learning;   5. Proposed or used sentiment analysis techniques.
Exclusion Criteria
1. Published in 2022;   2. Websites, leaflets, review, and survey literature;   3. Full-text not available online;   4. Paper is not associated with peer-reviewed;   5. Retracted and thesis paper;   6. Papers only rely on survey-based experiments;   7. Paper about only COVID-19.

2.1.2. Search String

A pilot search [31,32] was conducted on two renowned digital libraries, i.e., the institute for electrical and electronics engineers (IEEE) and the association for computing machinery (ACM), to determine the appropriate search string to ensure the generalizability of the findings. The search string consists of the keywords relevant to the *RQ*s. The purpose of this method is to determine relevant words or synonyms integrated into publications related to the sentiment analysis of public perception towards COVID-19 vaccination. We ran the pilot search several times, each time refining the search keywords in the search query. The search query was only implemented on the title and abstract of the publication. To prevent false positives, we decided to apply the search to a publication's metadata rather than the whole text. The final search string used in this study is shown below.

("Vaccine" OR "Vaccination") AND ("COVID-19" OR "Coronavirus" OR "Opinion Analysis" OR "Opinion Mining" OR "Sentiment Analysis" OR "Sentiment Classification" OR "Preception" OR "prospective" OR "Machine learning" OR "Artificial Intelligence " OR "predict" OR "Deep learning")

2.2. Execution

In this section, we discuss the procedure of processing and filtering the publications obtained from digital library searches. The initial pool of publications involving the retrieval of article information, i.e., title, abstract, publication year, and digital library name, yielded 1768 records. The "ScienceDirect" has the highest number of publications, i.e., 735. Following that, we employed a four-phase quality assessment procedure to exclude publications that do not fulfill our inclusion criteria. The volume of articles filtered at each step is depicted in Figure 1. As part of the quality assessment procedure, three authors manually reviewed the publications to determine whether a publication could progress from one phase to another. Phase-I begins with eliminating retracted and duplicate publications, which resulted in the removal of 151 publications. Phase-II involves screening the title and abstract of the remaining 1617 publications. A total of 1540 publications were excluded based on our inclusion and exclusion criteria. For instance, any article that did not adopt AI techniques or was not peer-reviewed was discarded. A total of 77 publications were retained, which were then subjected to a full-text scanning in phase-III, where the implication of our inclusion and exclusion criteria removed 40 publications, resulting in 37 publications as a starter set for the snowball sampling [33] in phase-IV. The snowball sampling fostered 10 more publications. In total, we identified 47 relevant publications: 37 publications in the first three phases and 10 in the snowball sampling.
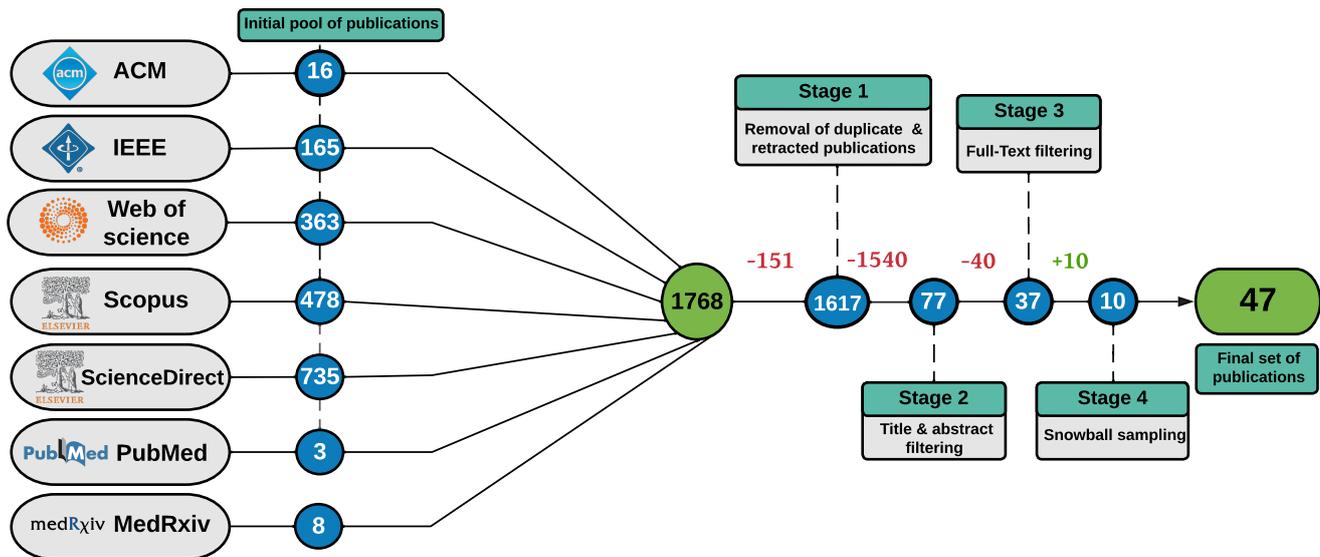
Figure 1. Overview of the number of publications resulting from the filtering process.

2.3. Synthesis

In this step, we synthesized the extracted data to answer the posed *RQs*. To begin answering *RQ₁*, we categorized the initial set of publications by looking at the algorithm on which sentiment analysis is based. For instance, publications utilizing machine learning algorithms, such as random forest (RF), Naive Bayes (NB), etc., are listed as learning-based studies. If an existing sentiment analysis tool such as TextBlob or valence aware dictionary for sentiment reasoning (VADER) is used in a study, we categorized it as a lexicon-based study. Some studies also utilized a unified framework of both aforementioned approaches and were listed as hybrid studies. The remaining publications that employed topic extraction tools, such as the latent Dirichlet algorithm (LDA), were classified as topic analysis studies.

In addition to this, we also cataloged the publications based on their year, month, and location; this assists in obtaining an overview of the characteristics of the extracted publications. Moreover, we analyzed the characteristics of the datasets (data source, number of records, availability status) utilized in the selected pool of publications to answer *RQ₂*. Another significant step is the thorough analysis of the public perception of the COVID-19 vaccine in accordance with the results of the selected publications. This helps us in answering our posed *RQ₃*.

Finally, to assure bias mitigation, all *RQ*-related data obtained during the publication review process were peer-reviewed, with discrepancies addressed through discussions. To help alleviate collaborative efforts throughout the author-review process, we used an Excel spreadsheet to record the manually extracted data. Moreover, this article uses a number of acronyms and abbreviations, which are listed in Section 6.

3. Results

This section presents the results for the designed *RQs* based on the analysis process detailed in Section 2.

3.1. RQ₁: Which Approaches Are Used in the Development of Sentiment Analysis Tools for Screening Public Perceptions of COVID-19 Vaccines?

To begin with, to answer *RQ₁*, we first present the statistics related to the sentiment analysis approaches employed in the selected studies. Secondly, we provide insights into the commonly used feature extraction techniques. Finally, we present a breakdown of the algorithms employed in the assessment of the public perception of the COVID-19 vaccines.

During a thorough analysis of the selected publications, we identified four approaches that are most widely utilized to analyze the public perceptions of the COVID-19 vaccines. Table 3 summarizes the approaches utilized in the finalized pool of publications.

3.1.1. Approach (1): Lexicon-Based Approach

The lexicon-based approach is an unsupervised (does not require a labeled dataset) sentiment screening of public opinions by incorporating a pre-defined sentiment lexicon [17]. A sentiment lexicon is a dictionary or a database comprised of opinion words tagged with their corresponding sentiments as positive, negative, or neutral [34]. A formula is used to aggregate the sentiment score of lexical words to predict the overall sentiment of an opinion. Studies that employed a lexicon-based approach are shown in Table 3.

3.1.2. Approach (2): Learning-Based Approach

The learning-based approach primarily employs word embedding to compute the vectorized score of the opinionated words [35]. It is a supervised sentiment analysis approach that incorporates labeled training data to learn predictive information in association with the target sentiments. The efficacy of the trained sentiment classifier is evaluated on the unlabeled test set. The ML classifiers combined with word embedding techniques used in the selected publications are listed in Table 3.

3.1.3. Approach (3): Hybrid Approach

The hybrid approach predominantly relies on the lexicon and learning-based sentiment analysis approaches to deal with unlabeled data [36]. This approach begins with the annotation of unlabeled data using a sentiment lexicon and moves forward with training and evaluating ML algorithms [37].

3.1.4. Approach (4): Topic Analysis

The topic analysis enables the probability estimation of the underlying text to each topic [38]. It groups the opinions based on the proximity corresponding to each term under consideration [39]. It also assists in determining the topics that are capturing more public attention. Based on the reviewed publications, the following four approaches were employed for the topic analysis:

- Latent Dirichlet allocation (LDA) is widely used in the field of topic modeling because of its effectiveness in identifying and extracting themes from a given corpus. As such, it can be applied to a document's text in order to determine what topic the text falls under. It creates models of both topics and words within documents using Dirichlet distributions.
- The Structural Topic Model (STM) is a generic framework for topic modeling that incorporates covariate data at the document level. It is permissible for the covariates to have an effect on either the topical prevalence, the topical content, or both. This can help enhance inference and qualitative interpretability.
- Linguistic Inquiry and Word Count (LIWC) count words by looking up their definitions in a dictionary organized by grammar, psychology, and topic. Because of its effectiveness in effectively classifying texts along psychological dimensions and predicting behavioral outcomes, LIWC has been widely applied as a text analysis tool across the social sciences.
- Correlation Explanation (CorEx) is a topic model that generates deep, informative topics from a collection of documents. The flexibility of CorEx to be performed as an unsupervised, semi-supervised, or hierarchical topic model makes it a compelling option compared to other topic models.

Among all four types of topic analysis, the most widely used topic analysis technique is LDA, which models the hidden topics by leveraging patterns related to similar phrases occurring consecutively and frequently together [40].

Table 3. Summary of approaches for screening of people's perception for COVID-19 vaccine

Ref.	Approach	Sentiment Lexicon	Features	Learning Model	Topic
[3]	Lexicon-based and Topic Analysis	EmoLex	N/A	×	LDA
[41]	Lexicon-based and Topic Analysis	TextBlob	N/A	×	LDA
[42]	Hybrid	Amazon Comprehend	TF-IDF	K-Means Clustering	×
[43]	Learning-based	×	TF-IDF	SVM, RF, LR, DT, KNN, GNB, AdaBoost	×
[44]	Learning-based	×	TF-IDF	SVM, KNN,	×
[45]	Learning-based	×	ANOVA	SVM, KNN, RF, ET, GBM, MLP, SGD, LR, DT, AdaBoot	×
[46]	Hybrid	VADER	TF-IDF	NB , LR	×
[47]	Learning-based	×	N/A	SDT, LR	×
[48]	Lexicon-based and Topic Analysis	CoreNLP	BoW	×	LDA
[49]	Learning-based	×	N/A	XGBoost, MLP, RF, K-Means Clustering	×
[50]	Learning-based	×	Doc2Vec, BOW	K-Means Clustering	×
[51]	Learning-based	×	N/A	NB, SVM, KNN, RF GB, AB, DT, LSTM, GRU, BERT, DNN	×
[52]	Lexicon-based and Topic Analysis	LIWC, VADER, Brandwatch	BoW	×	LDA
[53]	Lexicon-based and Topic Analysis	VADER	BoW	×	LDA
[54]	Lexicon-based and Topic Analysis	TextBlob	BoW	×	LDA
[55]	Learning-based and Topic Analysis	×	TF-IDF	BERT, LR, RF, SVM	LDA
[56]	Lexicon-based and Topic Analysis	TextBlob	BoW	×	LDA
[57]	Lexicon-based and Topic Analysis	VADER	BoW	×	LDA
[58]	Lexicon-based	VADER	BoW	×	×
[59]	Learning-based	×	Word2Vec	K-Means Clustering	×
[60]	Hybrid	ABSA	TF-IDF, Word2Vec	BERT	×
[61]	Hybrid	VADER	×	LSTM, Bi-LSTM	×
[62]	Hybrid	VADER, TextBlob	×	BERT	×
[63]	Hybrid	VADER, TextBlob	Word2vec	BERT	×
[64]	Learning-based	×	TF-IDF	BERT, Bi-LSTM, SVM, NB	×
[65]	Hybrid	TextBlob	×	KNN	×
[66]	Lexicon-based and Topic Analysis	SentiStrength	×	×	LIWC
[67]	Hybrid	TextBlob	BoW	NB	×
[68]	Hybrid	VADER	TF–IDF, N-grams (Uni, Bi, Tri)	LR, SVM, RF, DT, KNN, GNB	×
[69]	Learning-based	×	BoW	MNB, RF, SVM, Bi-LSTM, CNN, BERT	×
[70]	Hybrid	TextBlob	TF-IDF	SVM, RF, AdaBoost, MLP	×
[71]	Topic Analysis	×	N/A	×	LDA
[72]	Lexicon-based and Topic Analysis	TextBlob, VADER	N/A	×	LDA
[73]	Hybrid	TextBlob, VADER, ABSA	N/A	NB	×

Table 3. *Cont.*

Ref.	Approach	Sentiment Lexicon	Features	Learning Model	Topic
[74]	Learning-based	×	Word2Vec	CNN-LSTM	×
[75]	Lexicon-based	LIWC, TextMind	×	×	×
[76]	Learning-based and Topic Analysis	×	N/A	XLNet	LDA
[77]	Lexicon-based and Topic Analysis	VADER	N/A	×	LDA
[78]	Learning-based and Topic Analysis	×	TF-IDF	RF	LDA
[79]	Lexicon-based	SentimentR	×	×	×
[80]	Hybrid	VADER	BoW	SVM, KNN, LR, RF, M5, MLP, GPR	×
[81]	Lexicon-based and Topic Analysis	Amazon Comprehend	N-grams (Unigram, Bigram)	×	LDA
[82]	Topic Analysis	×	N/A	×	LDA
[83]	Learning-based	×	×	ANN	×
[84]	Lexicon-based	VADER	×	×	×
[85]	Learning-based and Topic Analysis	×	×	BERT	STM
[86]	Lexicon-based and Topic Analysis	VADER, TextBlob	N/A	×	CorEx

3.1.5. Findings for RQ_1

Of the total ($n = 47$) publications, the lexicon-based approach takes the lead with $\sim$36% ($n = 17$) studies, among which ($n = 13$) studies employed sentiment lexicons to perform topic analysis of the public opinions, whereas $\sim$32% ($n = 15$) belong to the learning-based, among which ($n = 4$) studies also performed topic analysis. The hybrid approach, on the other hand, comprises $\sim$26% ($n = 12$) of the filtered studies. The topic analysis approach is discretely adopted by only ($n = 2$) studies. However, if we combine the statistics, topic analysis has the highest rate of implementation in the selected pool of studies, which is $\sim$40% ($n = 19$). This reveals that many researchers attempted to exploit the main topics that were being discussed among people concerning the COVID-19 vaccine.

3.1.6. RQ_1-a: Which Sentiment Classifier Is Frequently Used?

In this part of RQ_1, we provide an overview of the sentiment classifiers utilized in the underlying studies. A sentiment classifier is a supervised machine learning or deep learning algorithm that is mainly used in learning-based and hybrid approaches [87]. It learns a function to map opinions (text data) to their respective (target) sentiments [88]. It infers a computation function from the labeled train data comprised of a set of data training samples. The performance of a sentiment classifier is evaluated using unseen test data containing a set of unlabeled testing samples. We provide a brief overview of the most popular algorithms used in analyzed studies:

- Random Forest (RF) employs a tree-based ensemble model. In this method, a number of smaller trees (decision trees) work together to make an accurate forecast [89]. Freestanding subtrees are created during training. Bagging is used to train these trees.
- Support Vector Machine (SVM) is used in classifiers and regression techniques [90]. SVM regression begins with the non-parametric technique, which is constructed on top of mathematical notation. The kernel transformation makes the input of desired data possible at this point. Regression issues are solved with the assistance of linear functions by the support vector machine.
- Naive Bayes (NB) is a controlled learning technique that is utilized to work through categorization issues [91]. This approach is predicated on the Bayes theorem. Because the training of an NB classifier only requires a small number of data points, the process is both quick and scalable. It is a probabilistic classifier that makes predictions of the likelihood of an object.
- Linear Regression (LR) is a statistical method that is frequently used by machine learning to carry out regression analysis [92]. It is possible to obtain an accurate picture of the connection that exists between the covariates or predictors, which are the independent variables, and the outcome variables, which are the dependent variables, by using this logical model.
- K-Nearest Neighbor (KNN) is a relatively straightforward model that may be implemented in ML to conduct analyses of regression and classification [93]. Using a distance function, the method sorts the new data into a class with its nearest neighbors, hence the name.
- Decision Tree (DT) can be thought of as a form of framework that resembles a tree and is used to construct constructions [94]. Due to the ease of use and the speed with which it may be carried out, decision trees are frequently utilized in the processing of medical information.
- BERT is fundamentally a transformer language model that can have any number of encoder layers and self-attention heads [95]. The goal of BERT is to provide computers with the ability to decipher the meaning of ambiguous material by analyzing the context of the surrounding text. The BERT framework was pre-trained with the help of text taken from Wikipedia, and it can be fine-tuned using question and answer datasets.

While investigating the filtered set of publications, we determined several sentiment classifiers integrated into the sentiment assessment of people's perception of the COVID-19 vaccine. Table 3 depicts that the most commonly used machine learning sentiment classifiers are RF ($n = 10$) and Support Vector Machine (SVM) ($n = 10$), followed by different types of NB ($n = 8$), such as Gaussian NB (GNB) and multinomial NB (MNB). In third place comes Logistic Regression (LR) ($n = 7$) and K-Nearest Neighbor (KNN) ($n = 7$), which are often chosen by researchers. Figure 2 depicts the distribution of machine learning sentiment classifiers used in the selected pool of publications.
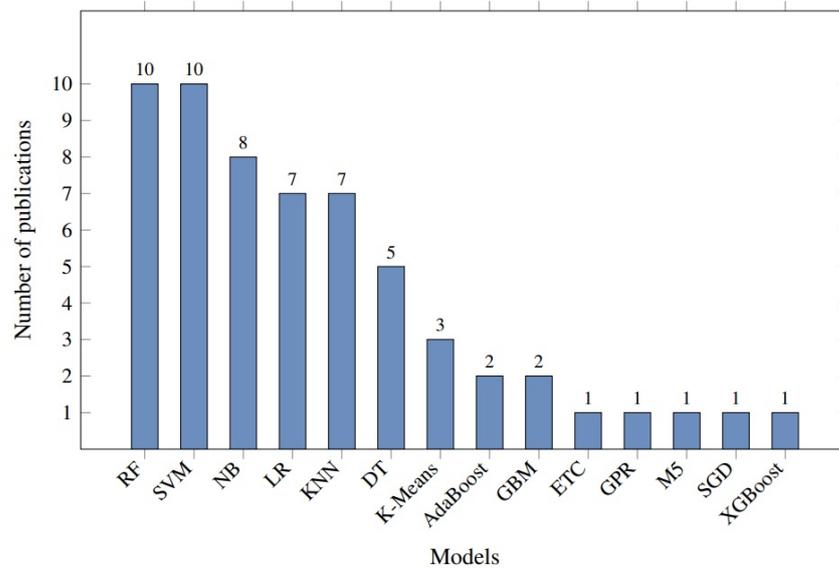

Figure 2. Distribution of machine learning-based sentiment classifiers used in learning-based approaches.

On the other hand, the frequently used deep learning sentiment classifiers are bidirectional encoder representations from transformers (BERT) ($n = 7$) and bidirectional long short-term memory (BiLSTM) ($n = 4$). Figure 3 presents the frequency of deep learning models used in the selected publications for this study.
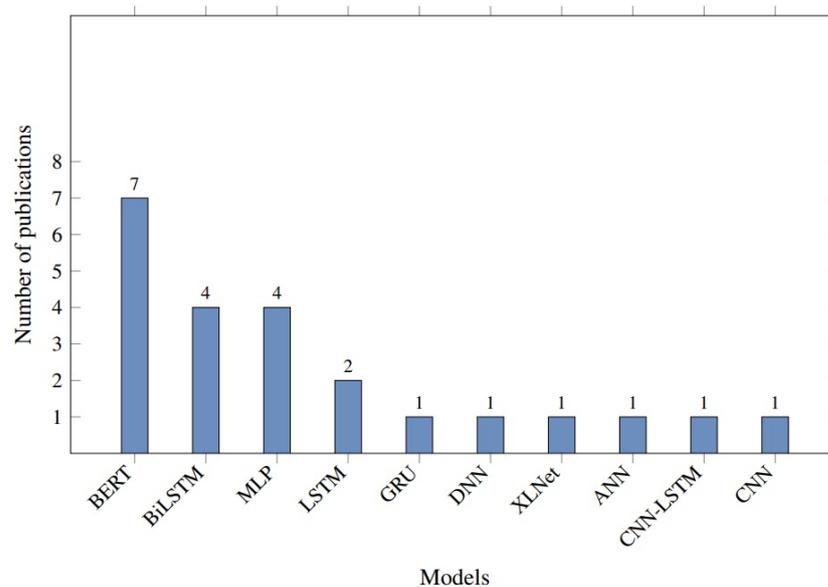

Figure 3. Distribution of deep learning sentiment classifiers used in learning-based approach.

Some studies also compared several learning algorithms. From the publications ($n = 47$), eight studies compared machine learning models and chose the best-performing algorithm. In these comparisons, SVM [43,44,70] and LR [43,46,68] stood out with highly accurate performances. When comparing deep learning models and machine learning models, BERT showed an outstanding performance [51,55,64,69], whereas in comparison to deep learning models in [61], BiLSTM yielded the best performance.

3.1.7. RQ_1-b: Which Feature Engineering Technique Has the Dominant Role in Sentiment Analysis?

Next, we investigate the feature engineering techniques employed in the available studies. Feature engineering or word embedding techniques are used to extract predictive information or features from the text data for effectively training the learning algorithms [96]. An effective word embedding technique can produce highly accurate results, as concluded by the authors of [97], which became the motivation for this part of RQ_1. We identified six feature engineering techniques that are extensively used in the selected publications, as shown in Table 4. We provide a brief overview of the six feature engineering techniques used in specific research:

- Term Frequency-Inverse Document Frequency (TF-IDF) is a statistical measure used to determine how important a word is to a set of documents. It is calculated by multiplying the inverse document frequency of a word across a group of documents by the number of times that word appears in a document.
- Bag of Words (BoW) is a classification technique that is frequently employed. It refers to a statistical language model that counts words to determine meaning in text or documents. The process of vectorization, which this approach performs, involves simply counting the number of times a word appears in a document. The end result is the text that is converted into vectors of a predetermined length.
- Word to Vector (Word2Vec) is a collection of different models used to create distributed representations of words found in a corpus. Word2Vec is an algorithm that takes a text corpus as an input and returns a vector representation of each word as an output.
- N-Grams refer to a string of n words or phrases that are consecutive inside a larger text or speech corpus. The N-Gram may consist of huge word clusters or smaller syllable groups. N-Grams are used as the foundation for operating N-Gram models, which play an important role in natural language processing as a means of forecasting incoming text or speech. N-Grams can be thought of as the building blocks of natural language.
- Document to Vector (Doc2Vec) is a Model that represents each Document as a Vector. The Doc2Vec model, in contrast to the Word2Vec model, is used to generate a vectorized representation of a document rather than a collection of words. It provides more information than just the average of the sentence's words.

Among all four types of topic analysis, the most widely used feature engineering techniques were determined to be Term Frequency-Inverse Document Frequency (TF-IDF) and Bag of Words (BoW). TF-IDF statistically assesses the relevance of a word in a document and BoW quantifies the occurrence of a word in a document irrespective of its significance in the document. The resulting feature sets of TF-IDF and BoW are often used by researchers to conduct sentiment analysis tasks.

The feature engineering technique is primarily utilized by machine learning models. Moreover, deep learning models perform automated feature engineering [98]; therefore, some studies [51,61,62] that employed deep learning models as sentiment classifiers did not utilize feature engineering techniques.

Table 4. Overview of the feature engineering techniques.

No.	Feature Engineering Technique	Total	Publications
1	BoW	10	[48,50,52–54,56,57,67,69,80]
2	TF-IDF	10	[42–44,46,55,60,64,68,70,78]
3	Word2Vec	4	[59,60,63,74]
4	N-Grams	2	[68,81]
5	ANOVA	1	[45]
6	Doc2Vec	1	[50]

3.1.8. Findings for RQ_1

The results reveal that there are four fundamental approaches used to understand the public's perception of the COVID-19 vaccines, including learning-based, lexicon-based, hybrid, and text analysis approaches. Out of 47 publications, 18 studies employed a topic analysis approach to investigate the main topics communicated among people during the pandemic. Conversely, the application of learning-based approaches and lexicon-based approaches have also been identified in sentiment screening of the public perception of COVID-19 vaccines. From the results, it can be seen that the lexicon-based approach is majorly used to annotate the unlabeled datasets, as manual annotation is a highly intensive task. Moreover, TF-IDF and BoW are predominantly used for vector representation of the text data (i.e., tweets, reviews, comments, etc.). Moving forward to the sentiment classifiers, a variety of machine learning algorithms and deep learning algorithms can be seen. However, SVM and RF, with their ability to map the relationship between text features and target variables with high efficacy, were identified as the recurring sentiment classifiers in the selected pool of publications.

3.2. RQ₂: What Are the Major Sources of Data to Monitor People's Opinions of COVID-19 Vaccines?

This *RQ* provides an overview of the data sources that were utilized to conduct sentiment analysis in the context of the COVID-19 vaccine. The majority of the datasets were acquired from mainstream social media platforms that have the largest user bases and application programming interfaces (APIs) that can handle large-scale data collection. Notably, data from Twitter were utilized by $n = 40$ publications from a total of 47 publications. Twitter has a high tendency to analyze and understand people's opinions regarding an entity [99,100]. Twitter is a micro-blogging online web forum that allows its users to publish short texts restricted to 280 characters, also known as Tweets.

The second major source for the collection of the public's opinions in the context of the COVID-19 vaccine is the questionnaire survey. From a total of 47 publications, $n = 5$ publications utilized online surveys to collect peoples' opinions and further categorize them based on their sentiments. Reddit, an online discussion forum that is well-known for its potential for interacting with different communities [101], was used by a total of two studies among the selected pool of publications to collect user posts in the underlying domain of the COVID-19 vaccine. Weibo, a Chinese micro-blogging platform [75], was also used to collect user posts. Facebook, another social media networking platform, was employed by researchers to collect Facebook users' posts regarding the COVID-19 vaccine [62]. Lastly, Google news, a well-known source of breaking news headlines, was used to collect news to perceive the public perception regarding the COVID-19 vaccine [63]. Figure 4 illustrates the frequency of each data source used in the finalized number of publications. Table 5 displays the data sources along with their characteristics and methods of collection. It also provides information regarding the number of records that were collected, along with the duration of data collection.
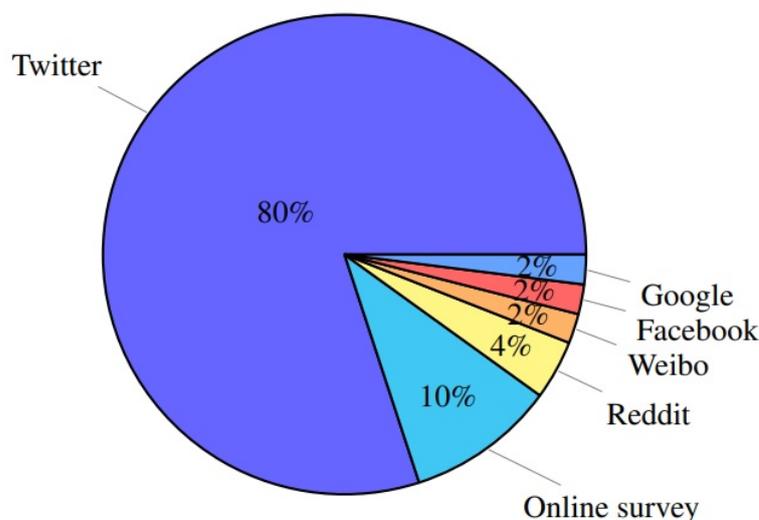
Figure 4. Frequency of data sources used in the publications for sentiment analysis in the context of COVID-19 vaccines.

3.2.1. RQ_2-a: What Is the Availability Status of Datasets Utilized to Asses the Public's Sentiments regarding COVID-19 Vaccines?

This part of RQ_2 covers the availability status of the datasets used in the underlying filtered studies. Dataset availability is of high significance as it helps novice researchers to perform their analyses. Answers to this question will help the readers by providing information regarding the datasets that are publicly available, along with the links from where they can access the datasets.

Among a total of 47 publications, only 16 publications made their dataset available for public use. Online surveys conducted by [49,84] were made publicly available. Out of 40 Twitter-based datasets, 12 are available, along with 1 Facebook, 2 Reddit, and 1 Google news datasets. Only two of the online surveys also provided the answers of their respondents for public use. The download links for each available dataset are provided in Table 5.

3.2.2. RQ_2-b: Which Techniques Are Employed to Annotate the Unlabeled Data Records According to Their Sentiments?

The user-generated data on online platforms are not labeled or tagged, whereas to understand the public perception of an entity, the data under consideration are required to be annotated with their respective sentiments. Regarding RQ_2-b, we provide an overview of the techniques employed in annotating the data records with their respective sentiments. From the reviewed publications, we identified two fundamental modes of sentiment tagging, including manual labeling and sentiment lexicon-based labeling. Manual labeling involves human annotators who read the text and then tag it with their respective sentiments, i.e., positive, negative, or neutral. It is a time-consuming, costly, and labor-intensive task. To address the aforementioned cons of manual labeling, sentiment lexicons are used to automate the sentiment tagging process.

Table 5. Characteristics of datasets employed to investigate the public's opinion of COVID-19 vaccines in the reviewed studies

Ref.	Source	Size	Duration	Tagging Technique	Availability
[3]	Twitter	1,499,421 Tweets	Mar 20–Jan 21	EmoLex	[102]
[41]	Reddit	1401 posts, 10,240 comments	Dec 20–May 21	TextBlob	[103]
[42]	Twitter	1 million Tweets	Mar 20–Jan 21	Amazon Comprehend	×
[43]	Twitter	14,022 Tweets	Feb 19–Mar 20	Manual	×
[44]	Twitter	5200 Tweets	×	Manual	×
[45]	Online Survey	175 Respondents	×	×	×
[46]	Twitter	1200 Tweets	×	VADER	×
[47]	Online Survey	6639 Respondents	Feb 21	×	×
[48]	Twitter	31,100 Tweets	Jan 20–Oct 20	CoreNLP	[104]
[49]	Online Survey	2237 Respondents	Apr 21	×	[105]
[50]	Twitter	12,134 Tweets	Jan 21	×	×
[51]	Online Survey	1647 Respondents	×	×	×
[52]	Twitter	185,953 Tweets	Nov 20–Feb 21	LIWC, VADER, Brandwatch	×
[53]	Twitter	902,138 Tweets	×	VADER	×
[54]	Twitter	154,978 Tweets	Mar 20–Aug 20	TextBlob	[106]
[55]	Twitter	5000 Tweets	Nov 20–Jan 21	Manual	×
[56]	Twitter	73,760 Tweets	Several months in 2021	TextBlob	×
[57]	Twitter	672,133 Tweets	Nov 20–Feb 21	VADER	[107]
[58]	Twitter	23,575 Tweets	Jan 21–Feb 21	VADER	×
[59]	Twitter	2,782,720 Tweets	Dec 19–Apr 21	×	×
[60]	Twitter	928,402 Tweets	Nov 20–Mar 21	ABSA	×
[61]	Twitter	125,906 Tweets	Dec 20–Nov 21	VADER	[108]
[62]	Facebook, Twitter	168,435 Facebook posts, 138,653 Tweets	Mar 20–Nov 20	VADER, TextBlob	[109]
[63]	Twitter, Google	637 Tweets, 569 Google news	Feb 20–May 20	VADER, TextBlob	[110]
[64]	Twitter	20,854 Tweets	Jan 20–Aug 20	×	[111]
[65]	Twitter	10,000 Tweets	×	TextBlob	×
[66]	Twitter	Over 12 million Tweets	Dec 20–Jan 21	SentiStrength	[112]
[67]	Twitter	190,000 Tweets	Dec 20–Jun 21	TextBlob	×
[68]	Twitter	431,986 Tweets	Feb 20–Apr 21	VADER	[113–115]
[69]	Twitter	752,951 Tweets	Nov 20–Dec 20	Manual	×
[70]	Twitter	13,109 Tweets	Jan 20–Jun 21	TextBlob	×
[71]	Twitter	2616 tweets	Dec 20–Jan 21	Manual	×
[72]	Twitter	980,557 Tweets	Nov 20–Dec 20	TextBlob, VADER	×
[73]	Twitter	21,000 Tweets	Feb 21–Mar 21	TextBlob, VADER, ABSA	×
[74]	Twitter	803,278 Tweets	Apr 21–Sep 21	LIWC, TextMind	×
[75]	Twitter, Weibo	756,118 Tweets, 362,950 Weibo posts	Dec 20–Feb 21	LIWC, TextMind	×
[76]	Twitter	4003 Tweets	Jul 20–Oct 20	Manual	×
[77]	Twitter	44,118 Tweets	Mar 20–Feb 21	VADER	×
[78]	Reddit	45,303 Comments	Jul 20–Jun 21	×	[116]

Table 5. *Cont.*

Ref.	Source	Size	Duration	Tagging Technique	Availability
[79]	Twitter	9036 Tweets	Feb 21–Mar 21	SentimentR	×
[80]	Twitter	899,663 Tweets	Apr 21–May 21	VADER	[117]
[81]	Twitter	144,101 Tweets	Aug 20–Jun 21	Amazon Comprehend	[118]
[82]	Twitter	100,209 Tweets	Feb 20–Mar 20	×	×
[83]	Online Survey	649 Respondents	Jan 21–Mar 21	×	[119]
[84]	Twitter	2.6 million Tweets	Nov 20–Jan 21	VADER	×
[85]	Twitter	16,959 Tweets	Mar 20–Jun 20	Manual	×
[86]	Twitter	2.4 million Tweets	Feb 20–Oct 20	VADER, TextBlob	×

The data is available in the Appendix of Ref. [104].

The majority of the studies ($n = 27$) utilized sentiment lexicons to annotate the data records, while ($n = 7$) performed manual labeling. However, some studies ($n = 13$) did not mention the method used for labeling the records of the collected dataset. The distribution of sentiment annotation techniques employed in the reviewed studies is shown in Figure 5. In the case of sentiment lexicons, the foremost used sentiment lexicon is VADER, which is a rule-based sentiment lexicon. It tags the text with their respective sentiments with an efficacy similar to a human annotator [17,120]. Another widely used sentiment lexicon in the reviewed studies is TextBlob, which computes the sentiment based on the subjectivity and polarity score of the underlying text [121,122]. Linguistic inquiry and word count (LIWC) was used by [53,75] to acquire the psycho-linguistic features [123] in the underlying public opinions. Amazon comprehend [124], a machine learning-based sentiment analyzer, is also used to tag the data records. TextMind is used to tag the sentiments of Chinese texts [125,126]. Aspect-based sentiment analysis (ABSA) [127], which extracts the aspect of the underlying text and categorizes the sentiment of each aspect, is also used by [60,73]. Another method to annotate the dataset used in the reviewed studies is SentimentR, which calculates the text polarity by assigning weights to polarity sifters, such as modifiers, negators, etc. [128]. Consequently, brandwatch [129], which was developed by Ph.D. qualifiers, incorporates knowledge-based, machine learning-based, and rule-based techniques to categorize the text as positive, negative, or neutral and is used by [52]. A variety of other sentiment lexicons such as CoreNLP [130], SentiStrength [131], and EmoLex [132] have also been used. The aforementioned techniques with their respective studies are provided in Table 5.
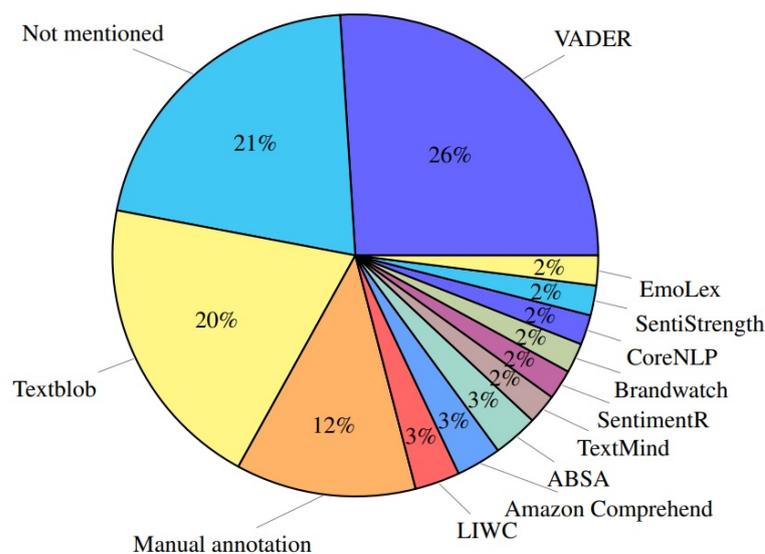
Figure 5. Distribution of sentiment annotation techniques.

3.2.3. Findings for RQ_2

The major source for analyzing the public perception of COVID-19 vaccines is determined to be Twitter. Moreover, a total of 16 studies made their datasets available for public use. The results reveal that researchers often employed VADER and TextBlob to tag their datasets as manual labeling is a time-consuming and costly task.

3.3. RQ_3: What Is the Public Perception of the COVID-19 Vaccine According to the Reviewed Studies Results? What Are the Geographical Locations Where Public Perception Has Been Studied regarding COVID-19 Vaccines?

COVID-19 vaccine-related views of the public were linked to different sentiments in different countries. To provide a detailed overview of public opinions of the COVID-19 vaccine, we stated the sentiment-related results of reviewed studies in Table 6. We also included "duration of underlying dataset", which corresponds to the date of data collection, to elaborate on the varying sentiments with respect to time in a particular country. The answer of RQ_3 will assist the readers in understanding the overall public perception of the COVID-19 vaccine in different countries.

The results of [48] indicated that several Australian Twitter users favored the COVID-19 vaccine and unsubstantiated misconceptions within the span of January 2020 to October 2020. Others who overlooked the dangers and seriousness of COVID-19 may have used conspiracy theories to justify their anti-vaccination stance. The study revealed three major topics of discussion regarding the COVID-19 vaccine, including "devising methods to control the infection", "fallacy and complaints", and "peoples' perception", with around 66.7% of tweets expressing positive sentiments and the remaining showing negative sentiments towards the COVID-19 vaccine. It also revealed eight emotions, among which "trust" yielded the highest percentage. However, according to the results revealed by [50], from January 2021, the prominent topics of discussion among Australian Twitter users shifted to vaccine roll-out, the willingness of the public to take the vaccine, the COVID-19 vaccine being a cause of death, and the approval of the COVID-19 vaccine.

Most of the studies that conducted sentiment analysis concerning the people of the USA demonstrated that the majority of the people were inclined towards either positive or neutral sentiments, whereas the ratio of negative sentiments remained the lowest among the people of USA [52,53,62,75,79,80]. This reveals that in the USA, people are more aware of vaccinating against infectious disease. The study [54] underscored the significance of five retrieved frequently discussed topics regarding the COVID-19 vaccine among US people. The authors found an overall negative sentiment concerning the "science" topic, whereas

other topics such as "coping without vaccine", "politics", "vaccine race", and "immunity boost" depicted positive sentiments. The study also underscored that negative tweets regarding vaccines were reacted to and retweeted more frequently. People who have had a bad personal experience with the COVID-19 outbreak are more likely to have negative thoughts about the vaccination, according to [76]. The public in the USA, on the other hand, is more focused on vaccination safety, effectiveness, and politics. The study [77] showed a growing trend in positive opinions in conjunction with the reduction in negative attitudes among Twitter users of the USA, reflecting an increase in anticipation and confidence in the COVID-19 vaccine. The public's emotions portray trust and anticipation for the vaccination, as well as a mixture of anger, fear, and sadness. Ref. [82] highlighted that Twitter users from the USA had a mixture of opinions regarding the COVID-19 vaccine, among which the news and those seeking information regarding COVID-19 and its vaccination constituted the majority of the discussion. Conspiracy theories highly contributed to the anti-vaccination discussions. With respect to the perception of people in India toward the COVID-19 vaccine, Ref. [56] states that a large part of Indian Twitter users have neutral sentiments, whereas a small portion of people shows negative sentiments towards the vaccination. Allergic reactions to vaccines and fear of death were also highlighted as two major topics of discussion among Indians.

The study [59] primarily focused on analyzing opinions of North American Twitter users on vaccination against influenza, followed by the COVID-19 vaccine. The findings revealed that the main focuses of discussion were three major topics, including health and medicine, health conditions, protection and responsibility, and politics. The first focuses on the tweets related to symptoms, the second studies the protection measures such as vaccination against the infection and the responsibility of taking the vaccine, while the third addresses the messages and opinions of the US government regarding their efforts to contain the infection.

The majority of Facebook and Twitter users in the United Kingdom (UK) expressed positive sentiments, with only 22% of the users expressing negative opinions towards the COVID-19 vaccine [62]. Moreover, the authors linked the positive sentiments to news on vaccine availability, vaccine development, and vaccine trials, whereas negative sentiments were more associated with the safety of the vaccine, delays in vaccine trials, and the availability of vaccines being affected by governments and companies.

Contrary to expectations, data from Twitter and Google news headlines regarding COVID-19 vaccines in Africa indicated more inclination towards positive sentiments over the period of February 2020 to May 2020 [63]. Similarly, on Weibo, the opinions of Chinese people concerning the COVID-19 vaccine have been more positive [75].

A study [74] investigating the sentiment orientation of Iranian people's opinions on homegrown and foreign vaccines revealed that the people of Iran were more predisposed to positive sentiments regarding foreign or imported vaccines.

Regarding Canadian Reddit users, an increasing trend in the discussion of vaccine supply and vaccine uptake is observed [78], while in Japan, neutral sentiments overwhelmed 85% of the tweets of Japanese users from August 2020 to June 2021, and negative sentiments dominated the remaining tweets [81].

Table 6. Public perception of COVID-19 vaccines according to reviewed studies.

Ref.	Country	Duration	Positive	Negative	Neutral	Topics
[3]	N/A	Mar 20–Jan 21	×	×	×	Emotions around vaccines (27.04%), knowledge around vaccines (23.7%), vaccines as a global issue (20.76%), vaccine administration (17.79%), progress on vaccine development and authorization (10.72%)
[41]	N/A	Dec 20–May 21	×	×	×	Vaccine, safety concerns, efficacy, potential side effects
[42]	N/A	Mar 20–Jan 21	7.2%	36.11%	55.11%	×
[48]	Australia	Jan 20–Oct 20	66.7%	33.3%	×	Devising methods to control the infection, Fallacy and complaints, Peoples' perception
[50]	Australia	Jan 21	×	×	×	Death, Approval, Hesitancy, Vaccine roll-out
[52]	USA	Nov 20–Feb 21	66.64%	33.64%	×	×
[53]	USA	×	48.2%	31.1%	20.7%	×
[54]	USA	Mar 20–Aug 20	×	×	×	Science (25.8%), Coping without vaccine (25.0%), Immunity boost (14.9%),Vaccine race (14.3%), Politics (20.0%).
[56]	India	Several months in 2021	36%	17%	47%	Fear over health, Rush in providing the vaccine, Allergic reactions, Various vaccines, Fear of death, Doubts regarding data, Skepticism over vaccine trails, Negative feeling towards pharma companies, Skepticism over the nationality of the vaccine, COVID-19 being exaggerated.
[57]	Worldwide	Nov 20–Feb 21	×	×	×	Emotional reactions (19.3%), public concerns (19.6%), news items (13.3%), public health (10.3%), vaccination drives (17.1%)
[58]	Worldwide	Jan 21–Feb 21	47.25%	18.75%	33.71%	×
[59]	North America	Dec 19–Apr 21	×	×	×	Health and medicine, Protection and responsibility, Politics
[62]	UK, USA	Mar 20–Nov 20	UK: 58%, USA: 56%	UK: 22%, USA: 24%	UK: 17%, USA: 18%	News regarding COVID-19 vaccine availability, Vaccine related trials, Vaccine development, Vaccine safety
[63]	Africa	Feb 20–May 20	51.02%	48.9%	×	×
[66]	Worldwide	Dec 20–Jan 21	×	×	×	Death, anger, and negative emotions
[71]	N/A	Dec 20–Jan 21	×	×	×	Misinformation/Conspiracy, Immunization Success, Mockery/Ridicule
[72]	N/A	Nov 20–Dec 20	×	×	×	Vaccine Performance, News and media coverage, People's aspirations, Companies and market, Healthcare environment, Vaccine potential and research

Table 6. *Cont.*

Ref.	Country	Duration	Positive	Negative	Neutral	Topics
[74]	Iran	Apr 21–Sep 21	43%	45%	12%	×
[75]	USA, China	Dec 20–Feb 21	USA: 30.62%, China: 40.64%	USA: 19.40%, China: 21.92%	USA: 49.99%, China: 37.44%	×
[76]	USA	Jul 20–Oct 20	×	×	×	Safety, effectiveness, and political issues regarding vaccines
[77]	USA	Mar 20–Feb 21	×	×	×	Trust, Anticipation, Fear, Sadness, Anger
[78]	Canada	Jul 20–Jun 21	×	×	×	Vaccine uptake, Vaccine supply
[79]	USA	Feb 21–Mar 21	35.30%	22.09%	36.40%	×
[80]	USA	Apr 21–May 21	25%	5%	70%	×
[81]	Japan	Aug 20–Jun 21	×	> positive	85%	×
[82]	USA	Feb 20–Mar 20	×	×	×	News Related to COVID-19 and Vaccine Development (26.2%), General Discussion and Seeking of Information on COVID-19 (25.4%), Financial Concerns (2.9%), Venting Negative Emotions (12.7%), Prayers and Calls for Positivity (9.9%), Efficacy of Vaccines and Treatments (8.1%), and Conspiracy theories (4.9%).
[84]	10 Countries	Nov 20–Jan 21	42.8%	30.3%	26.9%	×
[85]	N/A	Mar 20–Jun 20	×	×	×	Vaccine favorability, Vaccine unfavorability, Side Effects, Distrust, Conspiracy
[86]	N/A	Feb 20–Oct 20	×	×	×	fear (26.3%), anticipation (25.9%), trust (32.5%)

Public opinions from around the world showed temporal variation over time. The study [57] examined the main topics that were the center of discussion among Twitter users from November 2020 to February 2021. The results showed that the majority of tweets concerning emotional reactions and public concerns fall within the category of negative sentiments. Similarly, [58] showed that a large portion of extracted tweets from around the globe in the time span of one month, such as from January 2021 to February 2021, yielded positive sentiments, whereas a small portion of the tweets showed negative sentiments. Another study [66] showed that death, anger, and negative emotions are predominant themes being discussed on Twitter worldwide.

The authors of [42] did not specify any location; however, their findings highlighted an urgent need for proactive engagement with people from different educational backgrounds and cultures to validate the vaccine-related news and promote vaccine awareness since their study found that the second biggest group had a negative opinion regarding the COVID-19 vaccine. Other studies, such as [3,41,71,72,85] and [86], revealed that the public is more concerned with the safety and effectiveness of the COVID-19 vaccine, given the conspiracy theories circling on several social media platforms. This misinformation is misleading the public into emotions such as fear, anger, distrust, etc.

Regarding the geographical locations, we provided information on the geographical locations where sentiment analysis studies for the COVID-19 vaccine have been conducted. This will help readers and researchers who desire to expand this field of study identify what geographical areas are yet unexplored. The data for locations and their relevant reviewed papers are shown in Table 7. According to the statistics, the majority of the research focused on people's perceptions in the USA ($n = 12$). Countries with the highest frequency of COVID-19 cases, such as China, Italy, Iran, and Spain [133], were left behind in understanding the public opinion of vaccines. The locations of $n = 14$ studies were not mentioned. Studies that involved the UK public's ($n = 4$) and the worldwide public's ($n = 4$) perception contributed to a total of eight studies among the selected pool of reviewed studies.

Table 7. Geographical locations of reviewed studies

No.	Study Location	Total	Publications
1	Not Mentioned	14	[3,41–44,55,64,65,68,70–72,85,86]
2	United States	12	[52–54,60,62,67,75–77,79,80,82]
3	United Kingdom	4	[60,62,67,69]
4	India	3	[46,56,73]
5	World Wide	4	[57,58,61,66]
6	Australia	2	[48,50]
7	Canada	2	[60,78]
8	Africa	1	[63]
9	Japan	2	[67,81]
10	Qatar	1	[45]
11	22 Countries	1	[47]
12	Jordan	1	[49]
13	China	1	[75]
14	Iran	1	[74]
15	Portugal	1	[83]
16	10 Countries	1	[84]
17	Italy	1	[60]
18	Spain	1	[60]
19	Germany	1	[60]
20	France	1	[60]
21	Turkey	1	[60]
22	Bangladesh	1	[51]

Findings for RQ_3

The overall perception of people varies depending on their geographic region. Overall, neutral sentiments dominate in the reviewed studies, and positive sentiments topped the rest of the portion. People are increasingly concerned with vaccine development, safety, effectiveness, and adverse effects, according to the findings. People's negative emotions are strongly linked to conspiracy theories about the COVID-19 vaccination. Our findings show that the opinions of individuals in many other geographical locations on the COVID-19 vaccination have yet to be investigated.

4. Discussion

4.1. Challenges in Technology

One of our research questions was to identify the fundamental approaches used in understanding public perceptions, such as learning-based, text analysis, hybrid, and lexicon-based approaches, where we found that lexicon-based and text analysis approaches are the most commonly used. For vector presentation, the most popular approaches are TF-IDF and BoW. It would be prudent to state that when analyzing the collected data on vaccine hesitancy, technology posed a number of challenges to the data scientists. Data challenges included the nature of data, data collection, data annotation, and data processing.

4.1.1. Nature of Data

When dealing with data, we wanted to find the challenges that scholars faced when conducting sentiment analysis on the emotions and views of people regarding the uptake of vaccines. Challenges related to the nature of data mainly represent those aspects of data that hinder understanding the data. For instance, aspects such as the ambiguity of data and natural languages such as the use of irony, sarcasm, and slang language pose a challenge to data analysis [17,42,45,134]. Moreover, data noise in terms of grammatical errors and the use of incoherent text hinders a clear understanding of data [57,59]. Another challenge is the presence of unnecessary data [55,56,58,61]. In addition, different and multifaceted opinions on the same topic, as well as subjectivity and objectivity, may make it difficult to establish the nature of data [48,63,84,85].

4.1.2. Data Collection

We wanted to identify the major source of data for analyzing the public perception of the COVID-19 vaccine, which we found to be Twitter. Unfortunately, there are no studies that have used data from Instagram, Facebook, or other social media sites. It would be important to note that data collection represents a germane area when undertaking a sentiment analysis on any topic, particularly because the data are gathered from various social media outlets. However, data collection is often associated with a range of glitches. In this analysis, the identified issues were predominantly associated with the amount of data collected, data availability, location of data extraction, total duration for data collection, and language of the data. Challenges related to the volume of the collected data included issues of constructing a large dataset and concerns of small tweet sizes [52,83]. Challenges related to the availability of data include a change in online content through aspects such as deletion, strict character limitations, data access restrictions, and inconsistency of content [57,86]. With regards to the location of data extraction, previous studies have shown that geographical location is significant in understanding sentiment analysis [84,135]. Our study was consistent with such a finding, given that we noticed cross-country differences in their perceptions regarding COVID-19 vaccines. The majority of the studies were reviewed from the USA, followed by the UK and other areas. Such findings may inspire other researchers to replicate this study in other locations.

In our study, we acknowledged the fact that some of the reviewed countries do not have English as the official language. For example, countries in the Middle East majorly use Arabic, and others such as China use Chinese, among others. In our review, all studies performed their experiments in English, except a few that were conducted in Chinese,

Japanese, Hinglish (Hindi English), and Turkish. In previous studies, there has been language bias particularly for English tweets [43,79,86]. The duration of data collection was a challenge to researchers due to factors such as a short period of data collection and modification carried out on the time frame of data collection [64,80–82]. In addition to these challenges, there were also problems related to the cost of data collection and ethical considerations [82].

4.1.3. Data Annotation

After collecting data for sentiment analysis, the data must be synthesized through labeling to ensure clarity of the presented information. Challenges associated with annotation include laborious tasks, particularly when manually annotating a large volume of posts, and a lack of experts to offer guidance [65–68]. Moreover, annotation is challenging due to the strict demand for consistency and accuracy in data presentation [80]. In our study, we established that the labeling was predominantly conducted using VADER and TextBlob, which helped researchers overcome the aforementioned challenges. Such a finding may be very important for future researchers when undertaking data annotation.

4.1.4. Data Processing

In sentiment analysis, the stage after data collection and annotation is data processing. Data processing is challenging and problematic because it is carried out with different goals. In our study, one of our goals was to identify the most common machine learning sentiment classifiers, which included RF and SVM. We also found that the most popular deep learning sentiment classifiers used in the learning-based approach were BERT, followed by BiLSTM and multilayer perceptron (MLP). Such findings are very crucial because the major challenges in data processing are closely linked to techniques, inspection, and monitoring. Issues associated with processing techniques included the applicability of intelligent models such as deep learning or the time taken to process data using traditional processing techniques [62]. Moreover, there were issues related to the need to advance machine learning algorithms [64,75]. In data inspection and analysis, the identified challenges included the lack of adequate tools and the inability to identify the sentiment shared, particularly when sentiments were shared in the form of bots, abbreviations, hashtags, and emojis [54,56]. Finally, monitoring social media data was challenging as the researcher must have special skills to analyze the data [78].

4.1.5. Takeaway in Technology

We have provided a review of the available evidence and information regarding the perceptions of people toward COVID-19 and the techniques that have been used by previous researchers to achieve the aforementioned objective. Hence, our study takes a wide perspective and makes important implications for the technology and computer science field as follows:

- Our study has established which social media data are easy to acquire and utilize; some of it may be a bit difficult for scholars. Hence, we would like to indicate the need for social media sites to make their repositories and data more available.
- We also noted that several social media sites disabled the capacity to crawl their data. In this regard, we would like to encourage such platforms to allow information-crawling application programming interfaces (APIs) to assist researchers in perusing their data.
- It would also be prudent for us to recommend the development of additional technology to assist in data processing and analysis.
- We have also established that all the studies we used processed tweets in the English language tools because there are existing dictionaries, data sets, and analysis tools for that language. Hence, such a situation implies that the perspective we gave is majorly from English users.

- The authors also faced challenges with data labeling in cases of large data volumes, for which we recommend the use of VADER, and TextBlob, among others.

4.2. Social Challenges

Apart from data issues, we also identified a variety of social issues that were affecting public perceptions of COVID-19 vaccines. Social issues are those related to people, users, and society at large. The following sub-sections provide the social issues that we identified in our study.

4.2.1. Understanding

Our study identified various issues related to understanding vaccines that could affect public perceptions of them. For example, we noted that rumors [71], misconceptions [48], and incomplete/unclear information [71] could affect how people perceive vaccines. There is evidence to suggest that social understanding can be affected by the dynamic information on online platforms [82], information on social media, as well as other factors that may contribute to public sentiment. There has also been interest in the impact of online debates about vaccines, which have attracted attention from around the world [82], and the fact that such conversations target specific individuals and may lead to distortion of information [82].

4.2.2. Emotions

Our study revealed that both positive and negative emotions toward COVID-19 vaccines prevail. It would be plausible to suggest that emotions help in understanding how people think and live. Hence, they may also explain vaccine hesitancy. It has been found that emotions of fear [56], confusion [56], and lack of trust [77,85,86] could make people apprehensive about getting vaccinated.

4.2.3. Beliefs

There is a bi-directional relationship between the emotions of people and their beliefs, although it may be crucial to focus on beliefs separately. Our study showed that there are differing beliefs on COVID-19 vaccines, which could be classified as either individual beliefs or harmful information. Regarding individual beliefs, we found aspects such as conspiracy theories [71,82,85], beliefs [8], as well as notions of individual liberty, freedom, and responsibility [59]. On harmful information, we found that some people had distorted facts about vaccines [71,82], which they had taken from various online sources that resemble mainstream media [72]. It is important to state that although there has been published information discouraging people from taking vaccines, such ideas have been found to be erroneous [71,85] and are majorly blamed on social bots that seem to shape public views.

4.2.4. Behavior

Behavior involves the interplay between beliefs and emotions and could have important effects on people's attitudes and sentiments towards vaccines. People's behavior includes their online and offline activities. For instance, when people think that their online activity is monitored, they will desist from posting content, and if there is an attempt to correct their misinformation [71], there will be no desired effect if the intention is not aligned, which may not induce vaccine-associated behavior. With regards to the perspectives of people that are offline, some individuals tend to avoid getting vaccinated in some vaccine locations [72], which has often had negative implications on vaccination rates.

4.2.5. Strategies

We define strategies in this study as any undertakings that are made for the advantage or benefit of the population with respect to a given objective or goal. It would be prudent that a strategy is not limited to its understanding within the field of social sciences, and we look at it from the efforts that have been made to improve public perceptions of vaccines. The first strategy that has been used centers on coverage, which has been found

to be insufficient [72]. The second strategy is communication, which has raised concerns because it is poorly carried out [62]. The third strategy is related to promotions, which have been found to have insufficient information and are not persuasive [57]. Other strategies include countering fake news [48,71] and dealing with personal vaccine hesitancy aspects [50] and the inability to know the validity of the information. There were also issues in monitoring and difficulty in understanding how monitoring strategies affected vaccine perceptions [42]. Other challenges with monitoring included population bias, varied monitoring influencers [14], and the inability to focus on a specific period or across different times.

4.2.6. Takeaway in Social

We have established that the negative perceptions of COVID-19 vaccines are mostly attributed to misinformation and rumors that have been spread all over social media. It is evident that there is a lot of distrust, fear, and confusion that is preventing people from getting vaccinated. Our findings have the following implications:

- It is very important for authorities to take note of the misinformation that is going around on social media, especially among popular public figures that may shift public opinion. If necessary, influential people who spread fake COVID-19 vaccine information may need to be dealt with.
- We have also established that trust is very important for vaccine uptake. Therefore, public health authorities and healthcare workers must foster public trust in vaccines.
- The importance of communication has also been highlighted, especially when it comes from credible and reliable sources.
- From our research findings, we are also convinced that there is a need for a specialized measurement that will help in measuring vaccine hesitancy.
- Given the importance of social media sites, such as Twitter, in spreading information about COVID-19 vaccines, we would encourage public health authorities to partner with celebrities, artists, politicians, influencers, religious leaders, and other popular people in disseminating the right information.

4.3. Motivations

Motivations represent the driving forces that draw researchers and academics to a particular field of research. They demonstrate the benefits and value derived from pursuing a particular research endeavor. In this particular review that focuses on vaccine hesitancy and sentiment analysis, there are motivations related to technology as well as social factors. Motivations related to technology can be categorized into data availability, data accessibility, and usability of data.

4.3.1. Data Availability

Availability of data represents the value of having a large volume of data to analyze. Sentiment analysis is fundamentally an extensive area of research with plenty of available data, particularly on topics on health [50–54]. With a large amount of available data, researchers were drawn to the area to assess a large number of messages and large-scale information available [55].

4.3.2. Data Accessibility

Easily accessible data are advantageous to researchers in the domain of sentiment analysis. Researchers in this study were motivated by easy access to information. Moreover, technology made it possible to collect real-time data in a variety of languages.

4.3.3. Data Usability

Data usability explains the number of uses that a particular set of data has. Motivation for sentiment analysis from a technological perspective mainly included analysis, tracking, detection, and text extraction.

5. Recommendations

These recommendations mainly suggest strategies that can be adopted in order to improve research in sentiment analysis for other topics in the future.

5.1. Analysis Improvement

Recommendations for improving data analysis mainly address features, techniques, and datasets. These recommendations may assist researchers focusing on sentiment analysis to generate more knowledge from their chosen topics. In relation to features, it may be paramount to consider emojis and incorporate them as part of the sentiment analysis [16]. They may be integrated with machine-learning frameworks [136]. Moreover, more accurate algorithms for sentiment analysis can be developed and tested for topic extraction and classification of emotions [14,137]. In relation to data sets, it may be imperative to make use of larger datasets in analysis and data sources. Moreover, other social media tools other than Twitter can be used to collect data. Facebook, online message boards, and blogs are useful tools [136].

5.2. Social Recommendations

5.2.1. Vaccine

We have identified various aspects related to vaccines that may need attention. For example, it is important to utilize effective vaccine strategies that create positive public perceptions. There is also a need for studies investigating public sentiments towards vaccines and explanations of how social media information affects vaccine hesitancy.

5.2.2. Public

We have also identified useful strategies for getting the public to get vaccinated. For instance, we suggest offering vaccinations free of charge, utilizing influencers to encourage vaccine uptake, and investigating how popular beliefs affect Twitter followers. It is also important to handle individuals who spread fake information, explore what leads to vaccine information, and utilize social media to spread credible information to the public. Such efforts could significantly change public perceptions of vaccines.

6. Conclusions and Future Work

We conducted a comprehensive search of peer-reviewed scientific publications concerning the implementations of AI and ML techniques in monitoring the public's sentiments regarding the COVID-19 vaccine over a 2-year span (i.e., January 01, 2020, to December 31, 2021). This systematic review is an attempt to discuss the role of AI and ML in sentiment screening of public perception of the COVID-19 vaccine. The main challenges that a novice researcher encounters while conducting research are discussed. In particular, the area includes appropriate techniques, data sources, available datasets, and the latest work previously conducted in that domain, which is all explicitly addressed in this review study. Therefore, we consider this study as a one-stop source for the researchers attempting to broaden the research regarding the public's perception of the COVID-19 vaccine.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

TF-IDF	Term Frequency-Inverse Document Frequency
BoW	Bag of Words
ANOVA	Analysis of Variance
EmoLex	Word-Emotion Association Lexicon
Doc2Vec	Document to Vector
Word2Vec	Word to Vector
LIWC	Linguistic Inquiry and Word Count
ABSA	Aspect Based Sentiment Analysis
VADER	Valence Aware Dictionary and sEntiment Reasoner
CorEx	Correlation Explanation
STM	Structural Topic Model
LDA	Latent Dirichlet Allocation
SVM	Support Vector Machine
RF	Random Forest
LR	Linear Regression
DT	Decision Tree
KNN	K-Nearest Neighbour
GNB	Gaussian Naïve Bayes
AdaBoost	Adaptive Boosting
ET	Extra Trees
GBM	Gradient Boosting Algorithm
MLP	Multilayer Perceptron
SGD	Stochastic Gradient Descent
NB	Naive Bayes
SDT	Swinging Door Trending
XGBoost	Extreme Gradient Boosting
DNN	Deep Neural Network
GRU	Gated Recurrent Units
LSTM	Long Short-Term Memory
AB	Ada Boost
GB	Gradient Boosting
Bi-LSTM	Bidirectional Long Short Term Memory
CNN-LSTM	Convolutional Neural Network and Long Short-Term Memory Network
GPR	Gaussian Process Regression
ANN	Artificial Neural Network

References

1. World Health Organization. WHO Coronavirus (COVID-19) Dashboard. 2022. Available online: https://covid19.who.int/ (accessed on 27 July 2022).
2. Andreadakis, Z.; Kumar, A.; Román, R.G.; Tollefsen, S.; Saville, M.; Mayhew, S. The COVID-19 vaccine development landscape. *Nat. Rev. Drug Discov.* **2020**, *19*, 305–306.
3. Lyu, J.C.; Le Han, E.; Luli, G.K. COVID-19 vaccine–related discussion on Twitter: Topic modeling and sentiment analysis. *J. Med. Internet Res.* **2021**, *23*, e24435. [CrossRef] [PubMed]
4. MacDonald, N.E.; The SAGE Working Group on Vaccine Hesitancy. Vaccine hesitancy: Definition, scope and determinants. *Vaccine* **2015**, *33*, 4161–4164. [CrossRef] [PubMed]
5. Kwok, K.O.; Lai, F.; Wei, W.I.; Wong, S.Y.S.; Tang, J.W. Herd immunity–estimating the level required to halt the COVID-19 epidemics in affected countries. *J. Infect.* **2020**, *80*, e32–e33. [CrossRef]
6. Schuster, M.; Eskola, J.; Duclos, P.; The SAGE Working Group on Vaccine Hesitancy. Review of vaccine hesitancy: Rationale, remit and methods. *Vaccine* **2015**, *33*, 4157–4160. [CrossRef]
7. Phadke, V.K.; Bednarczyk, R.A.; Omer, S.B. Vaccine refusal and measles outbreaks in the US. *JAMA* **2020**, *324*, 1344–1345. [CrossRef]

8.   Wilder-Smith, A.B.; Qureshi, K. Resurgence of measles in Europe: A systematic review on parental attitudes and beliefs of measles vaccine. *J. Epidemiol. Glob. Health* **2020**, *10*, 46. [CrossRef]
9.   Khan, T.M.; Chiau, L.M. Polio vaccination in Pakistan: By force or by volition? *Lancet* **2015**, *386*, 1733. [CrossRef]
10. Onnela, J.P.; Landon, B.E.; Kahn, A.L.; Ahmed, D.; Verma, H.; O'Malley, A.J.; Bahl, S.; Sutter, R.W.; Christakis, N.A. Polio vaccine hesitancy in the networks and neighborhoods of Malegaon, India. *Soc. Sci. Med.* **2016**, *153*, 99–106. [CrossRef]
11. Taylor, S.; Khan, M.; Muhammad, A.; Akpala, O.; van Strien, M.; Morry, C.; Feek, W.; Ogden, E. Understanding vaccine hesitancy in polio eradication in northern Nigeria. *Vaccine* **2017**, *35*, 6438–6443. [CrossRef]
12. Shuto, M.; Kim, Y.; Okuyama, K.; Ouchi, K.; Ueichi, H.; Nnadi, C.; Larson, H.J.; Perez, G.; Sasaki, S. Understanding confidence in the human papillomavirus vaccine in Japan: A web-based survey of mothers, female adolescents, and healthcare professionals. *Hum. Vaccines Immunother.* **2021**, *17*, 3102–3112. [CrossRef]
13. Nehal, K.R.; Steendam, L.M.; Campos Ponce, M.; van der Hoeven, M.; Smit, G.S.A. Worldwide Vaccination Willingness for COVID-19: A Systematic Review and Meta-Analysis. *Vaccines* **2021**, *9*, 1071. [CrossRef]
14. Alamoodi, A.; Zaidan, B.; Al-Masawa, M.; Taresh, S.M.; Noman, S.; Ahmaro, I.Y.; Garfan, S.; Chen, J.; Ahmed, M.; Zaidan, A.; et al. Multi-perspectives systematic review on the applications of sentiment analysis for vaccine hesitancy. *Comput. Biol. Med.* **2021**, *139*, 104957. [CrossRef] [PubMed]
15. Joseph, C.B. Anti/Vax: Reframing the Vaccination Controversy. *J. Med. Libr. Assoc. JMLA* **2020**, *108*, 147. [CrossRef]
16. Piedrahita-Valdés, H.; Piedrahita-Castillo, D.; Bermejo-Higuera, J.; Guillem-Saiz, P.; Bermejo-Higuera, J.R.; Guillem-Saiz, J.; Sicilia-Montalvo, J.A.; Machío-Regidor, F. Vaccine hesitancy on social media: Sentiment analysis from June 2011 to April 2019. *Vaccines* **2021**, *9*, 28. [CrossRef] [PubMed]
17. Saad, E.; Din, S.; Jamil, R.; Rustam, F.; Mehmood, A.; Ashraf, I.; Choi, G.S. Determining the Efficiency of Drugs Under Special Conditions From Users' Reviews on Healthcare Web Forums. *IEEE Access* **2021**, *9*, 85721–85737. [CrossRef]
18. Rasheed, J.; Jamil, A.; Hameed, A.A.; Aftab, U.; Aftab, J.; Shah, S.A.; Draheim, D. A survey on artificial intelligence approaches in supporting frontline workers and decision makers for the COVID-19 pandemic. *Chaos Solitons Fractals* **2020**, *141*, 110337. [CrossRef]
19. Shah, F.M.; Joy, S.K.S.; Ahmed, F.; Hossain, T.; Humaira, M.; Ami, A.S.; Paul, S.; Jim, M.A.R.K.; Ahmed, S. A comprehensive survey of covid-19 detection using medical images. *SN Comput. Sci.* **2021**, *2*, 1–22. [CrossRef]
20. Lalmuanawma, S.; Hussain, J.; Chhakchhuak, L. Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review. *Chaos Solitons Fractals* **2020**, *139*, 110059. [CrossRef]
21. Ulhaq, A.; Khan, A.; Gomes, D.; Paul, M. Computer vision for COVID-19 control: A survey. *arXiv* **2020**, arXiv:2004.09420.
22. Agrawal, S. A survey on recent applications of Cloud computing in education: Covid-19 perspective. *J. Physics Conf. Ser.* **2021**, *1828*, 012076. [CrossRef]
23. Giri, B.; Pandey, S.; Shrestha, R.; Pokharel, K.; Ligler, F.S.; Neupane, B.B. Review of analytical performance of COVID-19 detection methods. *Anal. Bioanal. Chem.* **2021**, *413*, 35–48. [CrossRef] [PubMed]
24. Ashraf, I.; Alnumay, W.S.; Ali, R.; Hur, S.; Bashir, A.K.; Zikria, Y.B. Prediction Models for COVID-19 Integrating Age Groups, Gender, and Underlying Conditions. *Comput. Mater. Contin.* **2021**, *67*, 3009–3044. [CrossRef]
25. Aw, J.; Seng, J.J.B.; Seah, S.S.Y.; Low, L.L. COVID-19 vaccine hesitancy—A scoping review of literature in high-income countries. *Vaccines* **2021**, *9*, 900. [CrossRef]
26. Alamoodi, A.; Zaidan, B.B.; Zaidan, A.A.; Albahri, O.S.; Mohammed, K.; Malik, R.Q.; Almahdi, E.; Chyad, M.; Tareq, Z.; Albahri, A.S.; et al. Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: A systematic review. *Expert Syst. Appl.* **2021**, *167*, 114155. [CrossRef]
27. Petersen, K.; Feldt, R.; Mujtaba, S.; Mattsson, M. Systematic Mapping Studies in Software Engineering. In Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering (EASE), Bari, Italy, 26–27 June 2008.
28. Kitchenham, B.; Charters, S. Guidelines for Performing Systematic Literature Reviews in Software Engineering. 2007. Available online: https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.117.471 (accessed on 27 July 2022).
29. Aljedaani, W.; Peruma, A.; Aljohani, A.; Alotaibi, M.; Mkaouer, M.W.; Ouni, A.; Newman, C.D.; Ghallab, A.; Ludi, S. Test Smell Detection Tools: A Systematic Mapping Study. *Eval. Assess. Softw. Eng.* **2021**, 170–180. [CrossRef]
30. Aljedaani, W.; Krasniqi, R.; Aljedaani, S.; Mkaouer, M.W.; Ludi, S.; Al-Raddah, K. If online learning works for you, what about deaf students? Emerging challenges of online learning for deaf and hearing-impaired students during COVID-19: A literature review. *Univers. Access Inf. Soc.* **2022**, 1–20. [CrossRef]
31. Parkkila, J.; Ikonen, J.; Porras, J. Where is the research on connecting game worlds?—A systematic mapping study. *Comput. Sci. Rev.* **2015**, *18*, 46–58. [CrossRef]
32. Aljedaani, W.; Aljedaani, M.; AlOmar, E.A.; Mkaouer, M.W.; Ludi, S.; Khalaf, Y.B. I cannot see you—The perspectives of deaf students to online learning during covid-19 pandemic: Saudi arabia case study. *Educ. Sci.* **2021**, *11*, 712. [CrossRef]
33. Wohlin, C. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering, London, UK, 13–14 May 2014. [CrossRef]
34. Sallam, R.; Hussein, M.; Mousa, H. Improving collaborative filtering using lexicon-based sentiment analysis. *Int. J. Electr. Comput. Eng. (IJECE)* **2022**, *12*, 1744. [CrossRef]

35. Khoo, C.S.; Johnkhan, S.B. Lexicon-based sentiment analysis: Comparative evaluation of six sentiment lexicons. *J. Inf. Sci.* **2018**, *44*, 491–511. [CrossRef]

36. Chiny, M.; Chihab, M.; Chihab, Y.; Bencharel, O. LSTM, VADER and TF-IDF based Hybrid Sentiment Analysis Model. *Int. J. Adv. Comput. Sci. Appl. (IJACSA)* **2021**, *12*, 7. [CrossRef]

37. Mendon, S.; Dutta, P.; Behl, A.; Lessmann, S. A Hybrid Approach of Machine Learning and Lexicons to Sentiment Analysis: Enhanced Insights from Twitter Data of Natural Disasters. *Inf. Syst. Front.* **2021**, *23*, 1145–1168. [CrossRef]

38. Aljedaani, W.; Javed, Y.; Alenezi, M. LDA Categorization of Security Bug Reports in Chromium Projects. In Proceedings of the 2020 European Symposium on Software Engineering, Rome, Italy, 6–8 November 2020. [CrossRef]

39. Aljedaani, W.; Nagappan, M.; Adams, B.; Godfrey, M. A Comparison of Bugs Across the iOS and Android Platforms of Two Open Source Cross Platform Browser Apps. In Proceedings of the 2019 IEEE/ACM 6th International Conference on Mobile Software Engineering and Systems (MOBILESoft), Montreal, QC, Canada, 25–25 May 2019. [CrossRef]

40. Yu, H.; Yang, J. A direct LDA algorithm for high-dimensional data—with application to face recognition. *Pattern Recognit.* **2001**, *34*, 2067–2070. [CrossRef]

41. Melton, C.A.; Olusanya, O.A.; Ammar, N.; Shaban-Nejad, A. Public sentiment analysis and topic modeling regarding COVID-19 vaccines on the Reddit social media platform: A call to action for strengthening vaccine confidence. *J. Infect. Public Health* **2021**, *14*, 1505–1512. [CrossRef]

42. Elhishi, S.; El-Deeb, R.; El-Gamal, F.E.Z.A.; Sakr, N.A.; El-Metwally, S. Analyzing Public Perceptions Toward COVID-19 Vaccination Process Using Social Media and Machine Learning. In Proceedings of the 7th Annual International Conference on Arab Women in Computing in Conjunction with the 2nd Forum of Women in Research, Sharjah, United Arab Emirates, 25–26 August 2021; pp. 1–4.

43. Alhejaili, R.; Alhazmi, E.S.; Alsaeedi, A.; Yafooz, W.M. Sentiment Analysis of The Covid-19 Vaccine For Arabic Tweets Using Machine Learning. In Proceedings of the 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 3–4 September 2021; pp. 1–5.

44. Adamu, H.; Jiran, M.J.B.M.; Gan, K.H.; Samsudin, N.H. Text Analytics on Twitter Text-based Public Sentiment for Covid-19 Vaccine: A Machine Learning Approach. In Proceedings of the 2021 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET), Kota Kinabalu, Malaysia, 13–15 September 2021; pp. 1–6.

45. Hafizh, M.; Badri, Y.; Mahmud, S.; Hafez, A.; Choe, P. COVID-19 vaccine willingness and hesitancy among residents in Qatar: A quantitative analysis based on machine learning. *J. Hum. Behav. Soc. Environ.* **2021**, 1–24. [CrossRef]

46. Sivanantham, K. Sentiment Analysis on Social Media for Emotional Prediction During COVID-19 Pandemic Using Efficient Machine Learning Approach. *Comput. Intell. Healthc. Inform.* **2021**, 215–233. [CrossRef]

47. Riad, A.; Huang, Y.; Abdulqader, H.; Morgado, M.; Domnori, S.; Koščík, M.; Mendes, J.J.; Klugar, M.; Kateeb, E.; IADS-SCORE. Universal Predictors of Dental Students' Attitudes towards COVID-19 Vaccination: Machine Learning-Based Approach. *Vaccines* **2021**, *9*, 1158. [CrossRef]

48. Kwok, S.W.H.; Vadde, S.K.; Wang, G. Tweet topics and sentiments relating to COVID-19 vaccination among Australian Twitter users: Machine learning analysis. *J. Med. Internet Res.* **2021**, *23*, e26953. [CrossRef]

49. Hatmal, M.M.; Al-Hatamleh, M.A.; Olaimat, A.N.; Hatmal, M.; Alhaj-Qasem, D.M.; Olaimat, T.M.; Mohamud, R. Side Effects and Perceptions Following COVID-19 Vaccination in Jordan: A Randomized, Cross-Sectional Study Implementing Machine Learning for Predicting Severity of Side Effects. *Vaccines* **2021**, *9*, 556. [CrossRef]

50. Wang, G.; Kwok, S.W.H. Using K-Means Clustering Method with Doc2Vec to Understand the Twitter Users' Opinions on COVID-19 Vaccination. In Proceedings of the 2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI); Athens, Greece, 27–30 July 2021; pp. 1–4.

51. Chowdhury, A.A.; Das, A.; Saha, S.K.; Rahman, M.; Hasan, K.T. Sentiment Analysis of COVID-19 Vaccination from Survey Responses in Bangladesh. *J. Hum. Behav. Soc. Environ.* **2021** . [CrossRef]

52. Karami, A.; Zhu, M.; Goldschmidt, B.; Boyajieff, H.R.; Najafabadi, M.M. COVID-19 Vaccine and Social Media in the US: Exploring Emotions and Discussions on Twitter. *Vaccines* **2021**, *9*, 1059. [CrossRef] [PubMed]

53. Hung, M.; Lauren, E.; Hon, E.S.; Birmingham, W.C.; Xu, J.; Su, S.; Hon, S.D.; Park, J.; Dang, P.; Lipsky, M.S. Social network analysis of COVID-19 sentiments: Application of artificial intelligence. *J. Med. Internet Res.* **2020**, *22*, e22590. [CrossRef] [PubMed]

54. Wang, Y.; Shi, M.; Zhang, J. What public health campaigns can learn from people's Twitter reactions on mask-wearing and COVID-19 Vaccines: A topic modeling approach. *Cogent Soc. Sci.* **2021**, *7*, 1959728. [CrossRef]

55. Liu, S.; Li, J.; Liu, J. Leveraging Transfer Learning to Analyze Opinions, Attitudes, and Behavioral Intentions Toward COVID-19 Vaccines: Social Media Content and Temporal Analysis. *J. Med. Internet Res.* **2021**, *23*, e30251. [CrossRef]

56. Praveen, S.; Ittamalla, R.; Deepak, G. Analyzing the attitude of Indian citizens towards COVID-19 vaccine–A text analytics study. *Diabetes Metab. Syndr. Clin. Res. Rev.* **2021**, *15*, 595–599. [CrossRef]

57. Liew, T.M.; Lee, C.S. Examining the Utility of Social Media in COVID-19 Vaccination: Unsupervised Learning of 672,133 Twitter Posts. *JMIR Public Health Surveill.* **2021**, *7*, e29789. [CrossRef]

58. Mir, A.A.; Rathinam, S.; Gul, S. Public perception of COVID-19 vaccines from the digital footprints left on Twitter: Analyzing positive, neutral and negative sentiments of Twitterati. *Libr. Hi Tech* **2021**, *40*, 340–356. [CrossRef]

59. Benis, A.; Chatsubi, A.; Levner, E.; Ashkenazi, S. Change in Threads on Twitter Regarding Influenza, Vaccines, and Vaccination During the COVID-19 Pandemic: Artificial Intelligence–Based Infodemiology Study. *JMIR Infodemiol.* **2021**, *1*, e31983. [CrossRef]

60. Aygun, I.; Kaya, B.; Kaya, M. Aspect Based Twitter Sentiment Analysis on Vaccination and Vaccine Types in COVID-19 Pandemic With Deep Learning. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 2360–2369. [CrossRef]

61. Alam, K.N.; Khan, M.S.; Dhruba, A.R.; Khan, M.M.; Al-Amri, J.F.; Masud, M.; Rawashdeh, M. Deep Learning-Based Sentiment Analysis of COVID-19 Vaccination Responses from Twitter Data. *Comput. Math. Methods Med.* **2021**, *2021*, 4321131. [CrossRef] [PubMed]

62. Hussain, A.; Tahir, A.; Hussain, Z.; Sheikh, Z.; Gogate, M.; Dashtipour, K.; Ali, A.; Sheikh, A. 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*, e26627. [CrossRef]

63. Gbashi, S.; Adebo, O.A.; Doorsamy, W.; Njobeh, P.B. Systematic Delineation of Media Polarity on COVID-19 Vaccines in Africa: Computational Linguistic Modeling Study. *JMIR Med. Inform.* **2021**, *9*, e22916. [CrossRef] [PubMed]

64. To, Q.G.; To, K.G.; Huynh, V.A.N.; Nguyen, N.T.; Ngo, D.T.; Alley, S.J.; Tran, A.N.; Tran, A.N.; Pham, N.T.; Bui, T.X.; et al. Applying Machine Learning to Identify Anti-Vaccination Tweets during the COVID-19 Pandemic. *Int. J. Environ. Res. Public Health* **2021**, *18*, 4069. [CrossRef]

65. Shamrat, M.; Chakraborty, S.; Imran, M.; Muna, J.N.; Billah, M.M.; Das, P.; Rahman, O. Sentiment analysis on twitter tweets about COVID-19 vaccines using NLP and supervised KNN classification algorithm. *Indones. J. Electr. Eng. Comput. Sci.* **2021**, *23*, 463–470.

66. Malagoli, L.G.; Stancioli, J.; Ferreira, C.H.G.; Vasconcelos, M.; da Silva, A.P.C.; Almeida, J.M. A Look into COVID-19 Vaccination Debate on Twitter. In Proceedings of the 13th ACM Web Science Conference 2021, Virtual, 21–25 June 2021. [CrossRef]

67. Yang, X.; Sornlertlamvanich, V. Public Perception of COVID-19 Vaccine by Tweet Sentiment Analysis. In Proceedings of the 2021 International Electronics Symposium (IES), Surabaya, Indonesia, 29–30 September 2021. [CrossRef]

68. Jayasurya, G.G.; Kumar, S.; Singh, B.K.; Kumar, V. Analysis of Public Sentiment on COVID-19 Vaccination Using Twitter. *IEEE Trans. Comput. Soc. Syst.* **2022**, *9*, 1101–1111. [CrossRef]

69. Cotfas, L.A.; Delcea, C.; Roxin, I.; Ioanăş, C.; Gherai, D.S.; Tajariol, F. The Longest Month: Analyzing COVID-19 Vaccination Opinions Dynamics from Tweets in the Month following the First Vaccine Announcement. *IEEE Access* **2021**, *9*, 33203–33223. [CrossRef]

70. Amjad, A.; Qaiser, S.; Anwar, A.; ul Haq, I.; Ali, R. Analysing Public Sentiments Regarding COVID-19 Vaccines: A Sentiment Analysis Approach. In Proceedings of the 2021 IEEE International Smart Cities Conference (ISC2), Manchester, UK, 7–10 September 2021. [CrossRef]

71. Gokhale, S.S. Monitoring the Perception of Covid-19 Vaccine using Topic Models. In Proceedings of the 2020 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom), Exeter, UK, 17–19 December 2020; pp. 867–874.

72. Rahul, K.; Jindal, B.R.; Singh, K.; Meel, P. Analysing Public Sentiments Regarding COVID-19 Vaccine on Twitter. In Proceedings of the 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 19–20 March 2021; pp. 488–493.

73. Mudassir, M.A.; Mor, Y.; Munot, R.; Shankarmani, R. Sentiment Analysis of COVID-19 Vaccine Perception Using NLP. In Proceedings of the 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2–4 September 2021; pp. 516–521.

74. Nezhad, Z.B.; Deihimi, M.A. Twitter sentiment analysis from Iran about COVID 19 vaccine. *Diabetes Metab. Syndr. Clin. Res. Rev.* **2021**, *16*, 102367. [CrossRef]

75. Luo, C.; Chen, A.; Cui, B.; Liao, W. Exploring public perceptions of the COVID-19 vaccine online from a cultural perspective: Semantic network analysis of two social media platforms in the United States and China. *Telemat. Inform.* **2021**, *65*, 101712. [CrossRef]

76. Lyu, H.; Wang, J.; Wu, W.; Duong, V.; Zhang, X.; Dye, T.D.; Luo, J. Social media study of public opinions on potential COVID-19 vaccines: Informing dissent, disparities, and dissemination. *Intell. Med.* **2022**, *2*, 1–12. [CrossRef]

77. Hu, T.; Wang, S.; Luo, W.; Yan, Y.; Zhang, M.; Huang, X.; Liu, R.; Ly, K.; Kacker, V.; Li, Z. Revealing public opinion towards COVID-19 vaccines using Twitter data in the United States: A spatiotemporal perspective. *medRxiv* **2021**, 23, e30854.

78. Yan, C.; Law, M.; Nguyen, S.; Cheung, J.; Kong, J. Comparing Public Sentiment Toward COVID-19 Vaccines Across Canadian Cities: Analysis of Comments on Reddit. *J. Med. Internet Res.* **2021**, *23*, e32685. [CrossRef] [PubMed]

79. Ali, G.G.M.N.; Rahman, M.M.; Hossain, M.A.; Rahman, M.S.; Paul, K.C.; Thill, J.C.; Samuel, J. Public Perceptions about COVID-19 Vaccines: Policy Implications from US Spatiotemporal Sentiment Analytics. *Healthcare* **2021**, *9*, 1110. [CrossRef] [PubMed]

80. Sattar, N.S.; Arifuzzaman, S. COVID-19 Vaccination awareness and aftermath: Public sentiment analysis on Twitter data and vaccinated population prediction in the USA. *Appl. Sci.* **2021**, *11*, 6128. [CrossRef]

81. Niu, Q.; Liu, J.; Nagai-Tanima, M.; Aoyama, T.; Masaya, K.; Shinohara, Y.; Matsumura, N. Public Opinion and Sentiment Before and at the Beginning of COVID-19 Vaccinations in Japan: Twitter Analysis. *medRxiv* **2021**, *2*, e32335. [CrossRef]

82. Jiang, L.C.; Chu, T.H.; Sun, M. Characterization of Vaccine Tweets During the Early Stage of the COVID-19 Outbreak in the United States: Topic Modeling Analysis. *JMIR Infodemiol.* **2021**, *1*, e25636. [CrossRef]

83. Fernandes, N.; Costa, D.; Costa, D.; Keating, J.; Arantes, J. Predicting COVID-19 vaccination intention: The determinants of vaccine hesitancy. *Vaccines* **2021**, *9*, 1161. [CrossRef]

84. Liu, S.; Liu, J. Public attitudes toward COVID-19 vaccines on English-language Twitter: A sentiment analysis. *Vaccine* **2021**, *39*, 5499–5505. [CrossRef]

85. Jiang, X.; Su, M.H.; Hwang, J.; Lian, R.; Brauer, M.; Kim, S.; Shah, D. Polarization over vaccination: Ideological differences in Twitter expression about COVID-19 vaccine favorability and specific hesitancy concerns. *Soc. Media Soc.* **2021**, *7*, 20563051211048413. [CrossRef]

86. Saleh, S.N.; McDonald, S.A.; Basit, M.A.; Kumar, S.; Arasaratnam, R.J.; Perl, T.M.; Lehmann, C.U.; Medford, R.J. Public Perception of COVID-19 Vaccines through Analysis of Twitter Content and Users. *medRxiv* **2021**. [CrossRef]

87. Pandian, A.P. Performance Evaluation and Comparison using Deep Learning Techniques in Sentiment Analysis. *J. Soft Comput. Paradig. (JSCP)* **2021**, *3*, 123–134.

88. Rustam, F.; Reshi, A.A.; Aljedaani, W.; Alhossan, A.; Ishaq, A.; Shafi, S.; Lee, E.; Alrabiah, Z.; Alsuwailem, H.; Ahmad, A.; et al. Vector mosquito image classification using novel RIFS feature selection and machine learning models for disease epidemiology. *Saudi J. Biol. Sci.* **2022**, *29*, 583–594. [CrossRef] [PubMed]

89. AlOmar, E.A.; Aljedaani, W.; Tamjeed, M.; Mkaouer, M.W.; El-Glaly, Y.N. Finding the Needle in a Haystack: On the Automatic Identification of Accessibility User Reviews. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama Japan, 8–13 May 2021. [CrossRef]

90. Safdari, N.; Alrubaye, H.; Aljedaani, W.; Baez, B.B.; DiStasi, A.; Mkaouer, M.W. Learning to rank faulty source files for dependent bug reports. In *Big Data: Learning, Analytics, and Applications*; SPIE: Bellingham, WA, USA, 2019; Volume 10989, pp. 60–78.

91. Abid, M.A.; Ullah, S.; Siddique, M.A.; Mushtaq, M.F.; Aljedaani, W.; Rustam, F. Spam SMS filtering based on text features and supervised machine learning techniques. *Multimed. Tools Appl.* **2022.** [CrossRef]

92. Aljedaani, W.; Rustam, F.; Ludi, S.; Ouni, A.; Mkaouer, M.W. Learning sentiment analysis for accessibility user reviews. In Proceedings of the 2021 36th IEEE/ACM International Conference on Automated Software Engineering Workshops (ASEW), Melbourne, Australia, 15–19 November 2021; pp. 239–246.

93. Lee, E.; Rustam, F.; Washington, P.B.; El Barakaz, F.; Aljedaani, W.; Ashraf, I. Racism Detection by Analyzing Differential Opinions Through Sentiment Analysis of Tweets Using Stacked Ensemble GCR-NN Model. *IEEE Access* **2022**, *10*, 9717–9728. [CrossRef]

94. Aljedaani, W.; Mkaouer, M.W.; Ludi, S.; Ouni, A.; Jenhani, I. On the identification of accessibility bug reports in open source systems. In Proceedings of the 19th International Web for All Conference, Virtual, 25–26 April 2022. [CrossRef]

95. Adhikari, A.; Ram, A.; Tang, R.; Lin, J. Docbert: Bert for document classification. *arXiv* **2019**, arXiv:1904.08398.

96. Hakak, S.; Alazab, M.; Khan, S.; Gadekallu, T.R.; Maddikunta, P.K.R.; Khan, W.Z. An ensemble machine learning approach through effective feature extraction to classify fake news. *Future Gener. Comput. Syst.* **2021**, *117*, 47–58. [CrossRef]

97. Heaton, J. An empirical analysis of feature engineering for predictive modeling. In Proceedings of the SoutheastCon 2016, Norfolk, VA, USA, 30 March–3 April 2016; pp. 1–6.

98. Alissa, M.; Sim, K.; Hart, E. Algorithm selection using deep learning without feature extraction. In Proceedings of the Genetic and Evolutionary Computation Conference, Prague, Czech Republic, 13–17 July 2019. [CrossRef]

99. Pla, F.; Hurtado, L.F. Political tendency identification in twitter using sentiment analysis techniques. In Proceedings of the COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, Dublin, Ireland, 23–29 August 2014, pp. 183–192.

100. Pak, A.; Paroubek, P. Twitter as a corpus for sentiment analysis and opinion mining. In Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), Valletta, Malta, 19–21 May 2010; Volume 10, pp. 1320–1326.

101. Anderson, K.E. Ask me anything: What is Reddit? *Library Hi Tech News* **2015**, *32*, 8–11. [CrossRef]

102. Covid-19 Vaccine-Related Discussion on Twitter: Topic Modeling and Sentiment Analysis. Available online: https://zenodo.org/record/4104587 (accessed on 29 January 2022).

103. Public Sentiment Analysis and topic modeling regarding COVID-19 Vaccines on the Reddit Social Media Platform: A Call to Action for Strengthening Vaccine Confidence. Available online: https://github.com/Cheltone/NLP_Reddit (accessed on 29 January 2022).

104. Tweet Topics and Sentiments Relating to COVID-19 Vaccination among Australian Twitter Users: Machine Learning Analysis. Available online: https://www.jmir.org/2021/5/e26953/ (accessed on 1 February 2022).

105. Side Effects and Perceptions Following COVID-19 Vaccination in Jordan: A Randomized, Cross-Sectional Study Implementing Machine Learning for Predicting Severity of Side Effects. Available online: https://mdpi.com/article/10.3390/vaccines9060556/s1/ (accessed on 1 February 2022).

106. What Public Health Campaigns Can Learn from People's Twitter Reactions on Mask-Wearing and COVID-19 Vaccines: A Topic Modeling Approach. Available online: https://ieee-dataport.org/open-access/coronavirus-covid-19-tweets-dataset (accessed on 29 January 2022).

107. Examining the Utility of Social Media in COVID-19 Vaccination: Unsupervised Learning of 672,133 Twitter Posts. Available online: https://www.mdpi.com/2673-3986/2/3/24 (accessed on 29 January 2022).

108. Deep Learning-Based Sentiment Analysis of COVID-19 Vaccination Responses from Twitter Data. Available online: https://www.kaggle.com/gpreda/all-covid19-vaccines-tweets (accessed on 1 February 2022).

109. Artificial Intelligence Enabled Analysis of Public Attitudes on Facebook and Twitter toward COVID-19 Vaccines in the United Kingdom and the United States. Available online: https://gitlab.com/covid19aidashboard/covid-vaccination/ (accessed on 29 January 2022).

110. Systematic Delineation of Media Polarity on COVID-19 Vaccines in Africa: Computational Linguistic Modeling Study. Available online: https://medinform.jmir.org/2021/3/e22916/ (accessed on 1 February 2022).

111. Banda, J.M.; Tekumalla, R.; Wang, G.; Yu, J.; Liu, T.; Ding, Y.; Artemova, K.; Tutubalina, E.; Chowell, G. Applying Machine Learning to Identify Anti-Vaccination Tweets during the COVID-19 Pandemic. *arXiv* **2020**, arXiv:2004.03688.

112. A Look into COVID-19 Vaccination Debate on Twitter. Available online: https://zenodo.org/record/4721643#.YdUHBmjMLIV (accessed on 1 February 2022).

113. Analysis of Public Sentiment on COVID-19 Vaccination Using Twitter. Available online: https://www.kaggle.com/kaushiksuresh147/covidvaccine-tweets/version/24 (accessed on 29 January 2022).

114. Analysis of Public Sentiment on COVID-19 Vaccination Using Twitter. Available online: https://www.kaggle.com/ritesh2000 (accessed on 29 January 2022).

115. Analysis of Public Sentiment on COVID-19 Vaccination Using Twitter. Available online: https://www.kaggle.com/gpreda/all-covid19-vaccines-tweets/version/57 (accessed on 1 February 2022).

116. Comparing Public Sentiment Toward COVID-19 Vaccines Across Canadian Cities: Analysis of Comments on Reddit. Available online: https://www.openicpsr.org/openicpsr/project/120321/version/V12/view;jsessionid=509209814FC3AA43530A4E2C78C75FEA (accessed on 29 January 2022).

117. COVID-19 Vaccination Awareness and Aftermath: Public Sentiment Analysis on Twitter Data and Vaccinated Population Prediction in the USA. Available online: https://github.com/nawsafrin/covid-19 (accessed on 29 January 2022).

118. Public Opinion and Sentiment Before and at the Beginning of COVID-19 Vaccinations in Japan: Twitter Analysis. Available online: https://zenodo.org/record/3884334#.YwM3WXbMKUk (accessed on 29 January 2022).

119. Predicting COVID-19 Vaccination Intention: The Determinants of Vaccine Hesitancy. Available online: https://osf.io/tr2p3/ (accessed on 1 February 2022).

120. Hutto, C.; Gilbert, E. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the International AAAI Conference on Web and Social Media, Ann Arbor, MI, USA, 1–4 June 2014; Volume 8.

121. Khan, R.; Rustam, F.; Kanwal, K.; Mehmood, A.; Choi, G.S. US Based COVID-19 Tweets Sentiment Analysis Using TextBlob and Supervised Machine Learning Algorithms. In Proceedings of the 2021 International Conference on Artificial Intelligence (ICAI), Islamabad, Pakistan, 5–7 April 2021; pp. 1–8.

122. Hermansyah, R.; Sarno, R. Sentiment Analysis about Product and Service Evaluation of PT Telekomunikasi Indonesia Tbk from Tweets Using TextBlob, Naive Bayes & K-NN Method. In Proceedings of the 2020 International Seminar on Application for Technology of Information and Communication (iSemantic), Semarang, Indonesia, 19–20 September 2020; pp. 511–516.

123. Balage Filho, P.; Pardo, T.A.S.; Aluísio, S. An evaluation of the Brazilian Portuguese LIWC dictionary for sentiment analysis. In Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology, Fortaleza, Brazil, 21–23 October 2013.

124. Pinto, H.L.; Rocio, V. Combining Sentiment Analysis Scores to Improve Accuracy of Polarity Classification in MOOC Posts. In *Progress in Artificial Intelligence*; Springer International Publishing: Cham, Switzerland, 2019; pp. 35–46. [CrossRef]

125. Song, Y.; Kwon, K.H.; Lu, Y.; Fan, Y.; Li, B. The "Parallel Pandemic" in the Context of China: The Spread of Rumors and Rumor-Corrections During COVID-19 in Chinese Social Media. *Am. Behav. Sci.* **2021**, *65*, 2014–2036. [CrossRef]

126. Abe, M.; Aoki, K.; Ateniese, G.; Avanzi, R.; Beerliová, Z.; Billet, O.; Biryukov, A.; Blake, I.; Boyd, C.; Brier, E.; et al. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Preface. *Lect. Notes Comput. Sci. (Incl. Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinform.)* **2006**, *3960*, VI.

127. Tao, J.; Fang, X. Toward multi-label sentiment analysis: A transfer learning based approach. *J. Big Data* **2020**, *7*, 1–26. [CrossRef]

128. Naldi, M. A review of sentiment computation methods with R packages. *arXiv* **2019**, arXiv:1901.08319.

129. Pathak, A.R.; Agarwal, B.; Pandey, M.; Rautaray, S. Application of deep learning approaches for sentiment analysis. In *Deep Learning-Based Approaches for Sentiment Analysis*; Springer: Berlin, Germany, 2020; pp. 1–31.

130. Manning, C.D.; Surdeanu, M.; Bauer, J.; Finkel, J.R.; Bethard, S.; McClosky, D. The Stanford CoreNLP natural language processing toolkit. In Proceedings of the 52nd annual meeting of the association for computational linguistics: System demonstrations, Baltimore, MD, USA, 23–24 June 2014; pp. 55–60.

131. Thelwall, M. The Heart and soul of the web? Sentiment strength detection in the social web with SentiStrength. In *Cyberemotions*; Springer: Berlin, Germany, 2017; pp. 119–134.

132. Yavuz, M.C. Analyses of Character Emotions in Dramatic Works by Using EmoLex Unigrams. In Proceedings of the Seventh Italian Conference on Computational Linguistics CLiC-it 2020, Bologna, Italy, 1–3 March 2021; pp. 471–476. [CrossRef]

133. Miller, L.E.; Bhattacharyya, R.; Miller, A.L. Spatial analysis of global variability in Covid-19 burden. *Risk Manag. Healthc. Policy* **2020**, *13*, 519. [CrossRef] [PubMed]

134. Jamil, R.; Ashraf, I.; Rustam, F.; Saad, E.; Mehmood, A.; Choi, G.S. Detecting sarcasm in multi-domain datasets using convolutional neural networks and long short term memory network model. *PeerJ Comput. Sci.* **2021**, *7*, e645. [CrossRef]

135. Zhu, J.; Xu, C. Sina microblog sentiment in Beijing city parks as measure of demand for urban green space during the COVID-19. *Urban For. Urban Green.* **2021**, *58*, 126913. [CrossRef]

136. Yuan, X.; Schuchard, R.J.; Crooks, A.T. Examining emergent communities and social bots within the polarized online vaccination debate in Twitter. *Soc. Media Soc.* **2019**, *5*, 2056305119865465. [CrossRef]

137. Luo, X.; Zimet, G.; Shah, S. A natural language processing framework to analyse the opinions on HPV vaccination reflected in twitter over 10 years (2008–2017). *Hum. Vaccines Immunother.* **2019**, *15*, 1496–1504. [CrossRef]