Journal of Intelligent & Fuzzy Systems 45 (2023) 731–742
DOI:10.3233/JIFS-224086
IOS Press

Analyzing research trends of sentiment analysis and its applications for Coronavirus disease (COVID-19): A systematic review

Vipin jain* and Kanchan Lata Kashyap
VIT Bhopal University, Madhya Pradesh, India

Abstract. COVID-19 epidemic is one of the worst disaster which affected people worldwide. It has impacted whole civilization physically, monetarily, and also emotionally. Sentiment analysis is an important step to handle pandemic effectively. In this work, systematic literature review of sentiment analysis of Indian population towards COVID-19 and its vaccination is presented. Recent exiting works are considered from four primary databases including ACM, Web of Science, IEEE Explore, and Scopus. Total 40 publications from January 2020 to August 2022 are selected for systematic review after applying inclusion and exclusion algorithm. Existing works are analyzed in terms of various challenges encountered by the existing authors with collected datasets. It is analyzed that mainly three techniques namely lexical, machine and deep learning are used by various authors for sentiment analysis. Performance of various applied techniques are comparative analyzed. Direction of future research works with recommendations are highlighted.

Keywords: Sentiment analysis, COVID-19, opinion mining, neural networks, text classification

1. Introduction

SARS-CoV-2 spreaded a new global epidemic known as COVID-19 which horrified and shaken the entire globe [11, 18]. Millions of people infected by this harmful virus during pandemic. The economic growth of many countries also affected due to strict travel restrictions implemented by the governments [13]. Total 581,831,612 infected people and 6,413,423 fatalities recorded all over the world due to COVID-19 [66]. Total 564,126,546 confirmed and 6,371,354 death cases reported globally [28]. India is one of the most affected country by this. Total 43,847,065 confirmed with 525,930 death cases reported in India from March 2020 to June 2022. Figure 1(a) and (b) depicts the total confirmed and death cases of COVID-19 in India till 1 June 2022. COVID-19 transmits from person to person by coughing aerosols, droplets, and by contacting contaminated object or surface.

Enormous strain on healthcare infrastructure around the world has been seen due to this virus [5]. The scientific research community have also faced challenges to understand and trace behaviour of virus. Vaccination is one of the important and safest strategy to prevent this pandemic [2]. COVID-19 vaccination launched in India on 16 January 2021 for healthcare and front-line workers [22, 35]. After that, phase wise vaccination program has been conducted for common people. Mass vaccination is a challenging task due to growing anti-vaccination sentiments shared by the people. Emergence of social media allows users to express their opinions freely and openly [36]. Twitter is one of the social media used by many users to communicate with others by posting messages of 280 characters with emojis. Due to the severity of the COVID-19 outbreak, many people expressed their thoughts and feelings about this virus and vaccine on Twitter [68]. Many countries have implemented

*Corresponding author. Vipin jain, VIT Bhopal University, 466114, Madhya Pradesh, India. E-mail: vipin.jain2020@vitbhopal.ac.in.

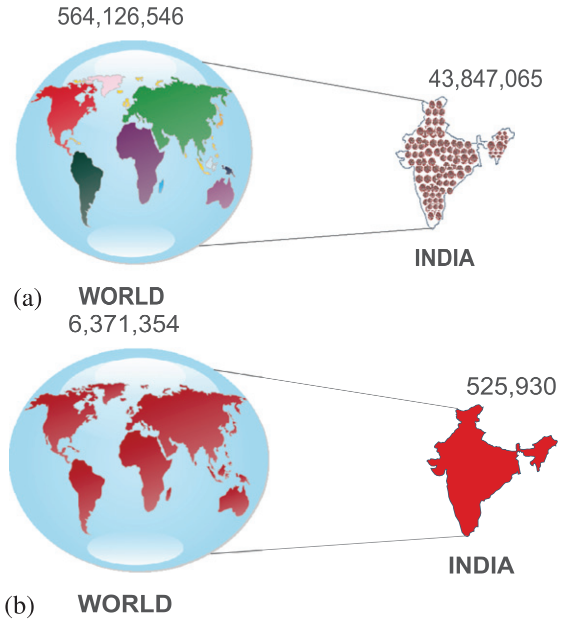

Fig. 1. (a) Confirmed cases (b) Dead cases.

total lockdown and stay home campaign. Users have shared their opinion about COVID-19 and it vaccine on Twitter during lockdown period. As a result, Twitter has become the most prominent medium to discuss about effect of COVID-19 and its vaccinations. Sentiment Analysis is the way to evaluate the emotions expressed by users [37]. Such analysis helps to divide the opinions into various categories such as positive, negative, and neutral [23]. Emotions can be categorized by applying various techniques such as (i) classical lexical method, (ii) machine learning, and (iii) deep-learning techniques. This study provides a comprehensive assessment of various techniques applied by exiting authors for sentiment analysis of COVID-19 and its vaccines.

1.1. Contributions

Numerous studies on COVID-19 sentiment classification have been conducted by various authors. This is the first review to examine Indian sentiments toward COVID-19 as well as its vaccination. The major contribution of this work is

- To collect the articles from numerous online sources and organise it based on the machine and deep learning models.
- To provide details of existing work in concise and understandable manner.
- To provide a classification and recommendation system for better understanding of existing work.

This study aims to identify data sources, data volume, and various techniques applied by exiting authors for sentiment analysis as well as future studies regarding COVID-19 and its vaccine. The rest of this paper is structured as follows: **Section 2** introduces the article selection process. The various methodologies applied in existing work are briefly described in **Section 3**. **Section 4** discusses the open issues and future research direction. Finally, **Section 6** concludes this work with various recommendations.

2. Capturing techniques of relevant articles

Complete online investigation is conducted with the peer-reviewed journals indexed in major online databases namely, web of science, IEEE, ACM, and Scopus. These digital libraries allow broad access of published articles. Systematic mapping technique is applied to examine the related articles of COVID-19 and its vaccination. Different steps are followed to collect relevant articles as shown in Fig. 2. It contains three steps (i) Article collection, (ii) Selection of article based on inclusion and exclusion criteria, and (iii) Screening of article identification and its eligibility. Figure 2 shows the screening procedure and inclusion criteria. Brief description about all steps are given in next sections.

2.1. Article collection

In this investigation, all COVID-19 and vaccination-related papers are retrieved from January 2020 to August 2022 without linguistic restrictions. Article search is started from scholarly repositories by using set of queries combined with AND and OR Boolean operators [26]. The sample of search queries are given as: "COVID-19” OR "COVID19” OR "coronavirus” OR "sars-cov-2” OR "SARS 2” AND "Sentiment Classification” AND "Sentiment Analysis ” AND ” Opinion Mining” OR "vaccine” OR "vaccination". Additional search criteria is narrowed by using publication year (2020, 2021, 2022), language, duplication and document type. Total 354 articles are collected after performing a literature search including 56 web of science, 97 IEEE, 143 Scopus, and 39 ACM.

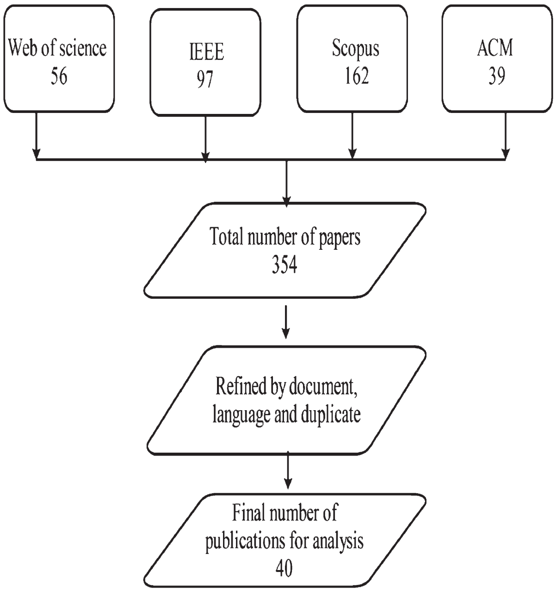

Fig. 2. Flow of data collection method.

2.2. Inclusion and exclusion criteria of papers

Further, most relevant papers are selected by adopting inclusion and exclusion parameters as shown in Table 1. Total 317 abstracts included for review after filtering out the duplicate articles.

2.3. Screening of article Identification and its eligibility

Article screening is done by (i) checking the titles and abstracts, and (ii) comprehensive study of selected articles. First, abstracts and titles of the remaining 317 papers are reviewed. After that, full-text reading is performed and created a final set of articles that fulfilled the inclusion criteria defined for this investigation. Finally, total 40 papers are included for analysis after excluding 277 publications.

3. Discussions and outcomes

Which approaches have been used for the development of COVID-19 and its vaccine sentiment analysis tools?

All collected articles are divided into three main categories: (i) first subcategory of articles covered lexicon based sentiment analysis, (ii) second group of articles addressed the machine learning-based emotion analysis, (iii) third group contains publications based on hybrid models of lexicon and machine learning techniques. Table 2 summarizes the COVID-19 tweet datasets used by the various authors for experimental analysis. Various techniques applied by the existing authors for sentiment analysis and classification are summarized in Table 3 and briefly described in next subsections.

3.1. Lexicon-based method

A lexicon based method for evaluating sentiments does not require human supervision [3, 54]. It determines the sentiment orientation based on the polarity value of the phrase [55]. Indian public opinion towards COVID-19 vaccination is determined by analyzing 73,760 tweets by Praveen et al. [49]. Authors have used topic modelling technique to grasp the difficulties of people with COVID-19 vaccination. They have have analyzed only 17% tweets are unfavourable while 47% are neutral. Many Indians are hesitant to get vaccinated due to fear of its adverse responses. Barkur et al. examined the Indian feelings about lockdown by using 24,000 tweets gathered between 25 March to 28 March 2020 [8]. Various unfavourable emotions such as anxiety, disgust, and sorrow towards lockdown are evaluated. A recent research used a multifaceted approach to explore the effect of COVID-19 epidemic on Indian emotions [48]. Authors analyzed positive attitudes of Indians toward lockdown and COVID-19. A web portal based on real-time tweet is proposed by Venigalla et al. to reflect the Indian sentiment during COVID-19 [62]. This platform allows visitors to check general sentiment of people of specific state on specific date and time.

Limitation of this study is that the current portal shows the state-level sentiment of few cities only.

Gupta et al. created a unique emotional care strategy to examine the heterogeneous linguistic data regarding COVID-19 [25]. Their study examined eight basic feelings across various topics such as ecology, security, healthcare, education, and economy. Twitter API is utilized to collect tweets from several regions of India. The published period of tweet is also chosen between time frame of 25 March to 09 June 2020. Only English tweets with the hashtags #COVID-19, #Lockdown, #LockdownDiaries, and #CoronaVirus are examined in their work. The emotion lexicon of National Research Council (NRC) is employed for sentiment evaluation which contains scores for various emotions such as anger, anticipation, trust, sorrow, pleasure, disgust, fear, and surprise. They

Table 1
Criteria for inclusion and exclusion used in this investigation

Inclusion Criteria	1.	Publication year(s): either 2020 or 2021
	2.	Written in English
	3.	Original research work
	4.	The article should propose or use artificial-based learning models.
	5.	The article should either suggest or use COVID-19 or its vaccination sentiment analysis methods.
exclusion criteria	1.	Not published between 2020 or 2021
	2.	Articles Written in Languages Other Than English
	3.	Referring to websites, conferences, reviews, book chapters, and literature surveys
	4.	Related to diseases other than COVID-19
	5.	Only COVID-19-related article

Table 2
Techniques adopted for opinion mining by Existing Authors in the literature

Ref	Lexicon	Learning Model
[62]	Textblob	HTML5
[59]	Minimum Redundancy Maximum Relevancy (mRMR)	BERT
[60]	Textblob	Ensemble Classifier
[39]	Textblob	SVM and Logistic Regression
[16]	–	BERT
[61]	Textblob	LDA
[6]	Textblob	LSTM
[1]	VADER	LDA
[9]	SentiStrength	BERT and GraphBERT
[44]	–	Random forest
[45]	R	LDA
[24]	TextBlob and VADER	Linear SVC classifier
[58]	Vader	–

have analyzed (9%-15%) reduction in happiness of people with 16% sad feelings and 18% scared about the health sector. Sv et al. investigated Indians perspectives regarding adverse effects of COVID-19 vaccines by using topic modelling technique [61]. Total 189,888 tweets containing "COVID Vaccine" and "Side effects" terms are collected by using Python Twint library. Authors have analyzed Indian emotions regarding risks associated with the COVID-19 vaccination. Total 78.5% of tweets are analyzed as neutral or favourable. Experiments indicated fear at workplace which leads to negative sentiments of Indians regarding COVID-19 vaccination. Rule based VEDAR technique is applied by by Mir et al. to analyze opinions of Indians on COVID-19 vaccination [43]. In addition, they identified common keywords used by the Indians to express their thoughts on Twitter regarding vaccinations. Total 11,815 Tweets are collected by using "Covid19vaccine" and "Coronavirusvaccine" hashtags from January to March 2021. Total 162 tweets are used for analysis after removing 2700 duplicate tweets. Total 639, 521, and 241 tweets are analyzed as positive, neutral, and negative, respectively. Most tweets are analyzed positive towards COVID vaccinations which indicates its widespread support. Dubey et al. examined Indian Twitter users' opinions about two COVID-19 vaccines namely, Covishield and Co-vaxin by using NRC lexicon [19]. They examined two datasets of tweets posted between January 14^{th} and January 18^{th}, 2021 with #Covishield and #Covaxin. Authors have analyzed favourable attitudes and trust of people towards both vaccines. Agarwal et al. experimented with news stories about internal migration in India during lockdown period due to COVID-19 [1]. They analyzed news articles published in Times of India and The Hindu newspapers. Total 2,170 separate news stories published between May 2020 and June 2020 are collected by using the terms "migrants," "migration," "lockdown," "COVID-19," and "pandemic". The tone

Table 3
Summary of dataset used by Existing Authors

Ref	Source	Size	Duration
[49]	Twitter	73,760	2020
[8]	Twitter	24,000	March 25^{th} to March 28^{th} 2020
[48]	Twitter	24,998	5^{th} April to 6^{th} April 2020
[20]	Instagram	–	April-May 2021
[1]	Times of India and The Hindu newspapers	2,170	May and June 2020
[62]	Twitter	194370	6^{th} May 2020
[59]	Twitter	596,784	20^{th} Jan to 25^{th} April 2020
[60]	Twitter	3100	23^{rd} March to 01^{th} Nov 2020
[25]	Twitter	8,84,111	25^{th} March to 09^{th} june 2020
[21]	Newspapers	100,000	January to December 2020
[16]	Github [50]	3090	23^{rd} March and 15^{th} July 2020
[61]	Twitter	189,888	March to April 2020
[43]	Twitter	11,815	4^{th} January to 22^{nd} March 2021
[6]	Twitter	10767	23^{rd} March to 02^{nd} April 2021
[1]	Twitter	2170	May and June 2020
[9]	Twitter	36,231,457	16^{th} January 2021 to 30^{th} November 2021
[44]	Twitter	48,913	March 2021 to June 2021
[45]	Twitter	50,000	23^{rd} March 2020 to 21^{th} May 2020
[34]	Sentiment 140 [4]	128,096	23^{rd} March 2020 to 13^{th} May 2020
[24]	Twitter	12,741	05^{th} April 2020 to 17^{th} April 2020
[58]	Twitter	401,037	03^{rd} May 2021 to 29^{th} August 2021

and perspective of news items are determined by using VADER module. Most articles are analyzed as neutral with small percentage of strong negative or positive polarity in their work. Chehal et al. examined emotions of Indians and their perception about e-commerce during both second and third lockdown periods using a Twitter dataset [15]. Less percentage of negative emotions are reported during third lockdown as compared to second lockdown period. Authors have analyzed the online retailing tendency of people. It is also analyzed that the priorities of people shifted from purchasing nutrition items, clothing, and home goods to baby goods, beauty aids, games, and sports equipment during third lockdown. Misra et al. examined and acquired information regarding reverse migration in India via Twitter mining [45]. They retrieved almost 50,000 Tweets from March 2020 to May 2020 by using trending hashtag such as #IndianMigrantWorkers and Twitter API. The emotions are identified by using the NRC Emotion Lexicon after noise removal from collected data.

Limitation: To obtain Twitter data, the researchers exclusively used only the popular hashtags #IndianMigrantWorker and #MigrantWorker which does not reflect the entire population. The tweets posted in other languages are not included for analysis. Different perspective can be analyzed by using tweets posted in other Indian languages also.

Sing et al. evaluated COVID-19 sentiment associated mucormycosis (CAM) during the second wave in India [58]. Total 401,037 Twitter posts are collected between 3^{rd} May and 29^{th} August 2021 by applying Twitter API. Higher percentage of positive emotion is analyzed as compared to negative emotions by using VADER tool.

3.2. Machine learning

The second category of research work addresses the application of machine learning techniques for the sentiment analysis of social media data regarding COVID-19 and its immunization. The machine learning techniques use the labelled training data to obtain predictive information regarding target opinions. Chintalapudi et al. collected Indian tweet posted between 23^{rd} March 2020 to 15^{th} July 2020 [16]. Collected data is categorised as happy, angry, scared, and sad by using Bi-directional Encoder Representation from Transformer (BERT) technique. Performance of BERT technique is compared with Long Short-Term Memory (LSTM), support vector machine (SVM), and logistic regression (LR) model also. Performance of BERT model is higher than other models with 89% accuracy. Kumar et al. applied hybrid model of BiLSTM and convolution neural network (CNN) to evaluate the publicly accessible Sentiment140 dataset and labelled Indian COVID-19 tweets and achieved

90% accuracy [34].
Limitation: Authors have used English text only for the sentiment analysis. The text from other languages can also be used to improve its correctness.

3.3. Hybrid models

A hybrid model combines lexical analysis and machine learning to analyze unlabeled data [17]. In this method, unlabeled data is annotated with lexicon-based algorithms before training and testing of machine learning algorithms [41].

Singh et al. investigated an emotion detection technique by using COVID-19 tweets of all over world [59]. Twitter scrapper API is used to extract data from 20 January to 25 April 2020 with #corona&virus, #COVID19, and #COVID2019 hashtags. Relevant features are selected with maximum Relevance and Minimum Redundancy (mRMR) technique. Sentiments are classified by utilizing BERT model with 94% accuracy. Sunitha et al. suggested an emotion analysis approach for evaluating real-time COVID-19 related tweets [60]. Approximately 3100 tweets are gathered between March 2020 to November 2021 from Indian and European citizens. Next, fasttext, Word2Vec, GloVe, and TF-IDF techniques are used for feature extraction from preprocessed data. The ensemble classifier is used to categorise the emotions as anger, sadness, joy, or fear. The suggested model successfully classified the emotions of both Indians and Europeans with an accuracy of 97.28% and 95.2%, respectively. Majumder et al. conducted comparative study of sentiment analysis using SVM and LR model [39]. The Indians COVID-19 tweets are gathered from March 2020 to June 2020. All collected data is converted into lowercase before removing hyperlinks and punctuation. Next, a label encoding approach is employed to get the labelled data by converting it into numeric format. Borah et al. employed a multi-modal deep learning approach to analyze 36,231,457 tweets related to COVID-19 vaccine from 51,682 Indians [9]. All Tweets are collected using #ReadyToVaccinate, #Covishield, #CovidVaccine, and #Covaxin hashtags. Analysis is done by using SentriStrength tools which assigns a value between -4 to +4 to each tweet. Extreme negative and extreme positive sentiment is denoted by -4 and +4, respectively. The textual data and the network topology are encoded by using BERT and GraphBERT.
Limitation: Sentiment analysis is applied only to tweets posted by urban residents. The perspective of rural residents is not included in the analysis. Furthermore, only limited hashtags determines inclusion of tweets in the dataset for analysis.
Gupta et al. evaluated 12,741 tweets to analyzed sentiments of Indian Twitter users by using natural language processing (NLP) and machine learning techniques [24]. Linear SVC classifier with unigrams is used for classification and achieved maximum 84.4% accuracy. Positive attitude of Indians towards lockdown decision of government is analyzed.

Ghasiya et al. identified COVID-19 relevant issues and sentiments published in newspapers between January to December 2020 [21]. Total 100,000 news articles are scraped with COVID-19 and Coronavirus keywords from eight major newspapers of four countries. Their work is divided into two parts (i) topic modification and (ii) sentiment classification. According to topic modelling, all four countries have similar issues of sports, education, and economy. After that, they used state-of-the-art RoBERTa model to determine sentiment of headlines and achieved 90% validation accuracy. Their findings indicate more positive news in South Korea and UK as compared to negative.

Xie et al. investigated reaction of Chinese microblog users about COVID-19 using text mining techniques [67]. They collected web crawler of 719,570 Weibo posts. It is analyzed that people supported the front-line soldiers during COVID-19 outbreak and positive messages percentage dominated negative messages.

Ermatita et al. proposed a multi-modal fusion neural network for COVID-19 sentiment analysis of Instagram text and posters [20]. Integrated inputs of images and captions are given to modified deep learning architectures with multi-modal graph layers and self-attention. Authors obtained 87% accuracy with multi-modal Fusion Neural Networks.

Gupta et al. proposed a novel emotional care scheme for analysing real-time COVID-19 textual data [25]. They have analyzed eight emotions towards various categories such as politics, market, education, health, lockdown, and nature. According to this textual analysis, 'joy' feeling reported less towards everything (9-15%) except nature (17%).

3.4. Sentiment analysis approaches

Various techniques applied by the exiting authors for COVID-19 sentiment analysis are briefly described in following sections.

Table 4
Comparison of existing models

Ref	Model	Accuracy (%)
[57]	LSTM-RNN	84.56
[23]	Linear SVC	98.15
[42]	CNN-Bi-LSTM	99.33
[31]	RNN	93.02
[56]	Naive Bayes	91
[30]	(H-SVM)	96
[53]	ETC	93
[29]	LSTM	81.15
[14]	LogisticRegression	81
[27]	LSTM+FastText	82.4
[65]	BERT	75.65
[53]	LSTM	93

3.4.1. Lexicons based sentiment analysis

Lexicons are collections of words or terms used for sentiment analysis in natural language processing. It can be created manually or extracted automatically from a text corpus. Lexicon contains both positive and negative words or phrases with polarity scores. Several sentiment based lexicons resources are described as: (i) SentiWordNet: It is a lexical resource that assigns sentiment scores to each word based on their senses and WordNet.
(ii) VADER (Valence Aware Dictionary and sEntiment Reasoner): It is a a rule-based sentiment analysis tool that uses a lexicon of words and emoticons to determine text sentiment.
(iii) TextBlob: It is a Python based simple API which provides part-of-speech tagging and noun phrase extraction. This method is based on modified version of Naive Bayes algorithm which provides sentiment polarity score between -1 and +1. Highly negative, highly positive, and neutral sentiment are denoted by +1, -1, and 0, respectively. Polarity score can be computed as:

$$P(S|W) = P(W|S) * P(S)/P(W) \tag{1}$$

$P(S|W)$ is probability of sentiment S with given words W. $P(W|S)$ is probability of words W with given sentiment S. $P(S)$ denotes prior probability of sentiment S. $P(W)$ indicates probability of words W.

3.4.2. Logistic regression

It is statistical based categorization model that assesses correlation between categorical dependent and one or more independent variables [52]. Proper feature selection can improve the accuracy and generalizability of model. Mathematical expression of logistic regression can be given as:

$$p(y = 1|x, w) = 1/(1 + e^{-(wx+b)}) \tag{2}$$

here, $p(y = 1|x, w)$ denotes the predicted probability of positive class with given input features x and model weights w. Exponential function and bias term are denoted as e and b, respectively.

3.4.3. Naive Bayes

It is probabilistic-based classifier which estimates likelihood of a group [40]. Only small quantity of data is needed to train this classifier [47]. It produces better outcome because of its simplicity and stable foundation. Bayes theorem is mathematically expressed as [33, 38]:

$$C(F \mid R) = \frac{C(R \mid F) \cdot C(F)}{C(R)} \tag{3}$$

here, $C(F \mid R)$ denotes class probability F of given document R. $C(F)$ represents prior probability of class F, $C(R)$ denotes knowledge from the text itself to be categorised. $C(R \mid F)$ represents document probability R having distribution in class space F.

3.4.4. Random forest classifier

This classifier builds a forest of decision trees to handle complex classification and regression problems [10]. It generates numerous decision-making models during training to predict different classes. It employs Gini Index and Entropy for data categorization which are mathematically represented as:

$$\text{Gini Impurity } = 1 - \sum_{i=1}^{k} p_i^2 \tag{4}$$

$$\text{Entropy } = - \sum_{i=1}^{k} p_i \log_2 p_i \tag{5}$$

here, k and p_i denotes number of classes and proportion of samples belonging to i^{th} class, respectively. Random Forest constructs decision trees by minimizing the impurity measure at each node to accurately classify the new instances.

3.4.5. Convolution neural network (CNN)

It is a feed-forward neural network which contains four components (i) hidden layers, (ii) convolution layer, (iii) ReLU layer, and (iv) pooling layer to extract data-based features [46]. ReLU and pooling layers are standard components which uses grid-like structure to extracts essential features from given input data. All features are generated by using a window of words x_t and it is mathematically given as:

$$S = T(Q \cdot x_t + p) \tag{6}$$

here, filter weight and bias is denoted by Q and p, respectively. The T presents convolutional nonlinear activation function.

3.4.6. LSTM

It is a type of recurrent neural network (RNN) commonly used for text classification. It uses a combination of three main gates: (i)input, (ii) forget, and (iii) output, to control the flow of information. Input gates control the amount of information added to memory cells at each time step. Forget gate determines amount of time for which memory cell retains the old information. Finally, output gate controls the amount of current memory cell information used to generate output at each time step. All three gates are mathematically denoted as: [12, 63]:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \tag{7}$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \tag{8}$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \tag{9}$$

here, x_t is given input at time step t, h_{t-1} is the hidden state from the previous time step. Input, forget, and output gate activation at time step t are denoted as i_t, f_t, and o_t, respectively. σ denotes the sigmoid function.

3.4.7. BiLSTM

It is variation of the LSTM network that integrates both past and future inputs in a single time step. It utilises bidirectional LSTM layer to uncover discoverable patterns by traversing the input data history in both directions. LSTM classifier works well with variable-length sequences but unable to use contextual information from future tokens [51]. First and second layer of BiLSTM network traverses the text in forward and reverse sequence, respectively [32]. Finally, output layer deals with historical and prospective context of each sequence point. The bidirectional architecture of BiLSTM improves its capability to understand the meaning of given text. The mathematical equation of this model is given as:

$$\overrightarrow{h_t} = LSTM(\overrightarrow{h_{t-1}}, x_t) \tag{10}$$

$$\overleftarrow{h_t} = LSTM(\overleftarrow{h_{t+1}}, x_t) \tag{11}$$

$$h_t = [\overrightarrow{h_t}; \overleftarrow{h_t}] \tag{12}$$

$$y = softmax(W_{hy}h_t + b_y) \tag{13}$$

here, $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$ denotes hidden states of forward and backward LSTMs at time t, respectively. LSTM input at time t is denoted as x_t. Concatenation of both hidden states are represented by [;]. Weight matrix and bias vector for fully connected layer are represented by W_{hy} and b_y. Predicted class probability is given as y.

3.4.8. Gated recurrent unit (GRU)

This model regulates the internal flow of information with a gating mechanism namely update and reset gates [64]. Both gates determine the amount of data accepted or discarded from the previous level. Mathematically, both reset gate and update gate are defined as [69]:

$$q_a = \sigma(T_q * [p_{a-1}, m_a]) \tag{14}$$

$$l_a = \sigma(T_l * [p_{a-1}, m_a]) \tag{15}$$

here, q_a and l_a are reset and update gates at time step a, respectively; p_{a-1} is the hidden state at the previous time step; m_a is the input at time step a; T_q and T_l are weight matrices to be learned during training. Output of this model at time step a is calculated as:

$$p_a = (1 - l_a) * p_{a-1} + l_a * tanh(T_p * [q_a * p_{a-1}, m_a]) \tag{16}$$

here, T_p is another weight matrix to be learned during training. Output p_a is passed through a final Softmax layer to obtain predicted class probabilities [7].

4. Discussions

Two significant aspects is examined in this work listed as: (i) Various challenges encountered by existing researchers. (ii) Relevance and benefits of COVID-19 sentiment analysis. Various challenges, motivations, and recommendations are briefly described in next sections.

4.1. Challenges

Existing researchers have faced several technological challenges while assessing the COVID-19 vaccines data. The data type, its annotation, and data pre-processing are the main challenges faced by the researchers. Accurate data plays a key role to perform accurate analysis and to get valid conclusion. It is difficult to accurately analyze data with irony, sarcasm, and slang words by using natural language processing. Credibility and originality of collected data are two crucial challenges of sentiment analysis on social media. Regrettably, only few authors have utilized data from Instagram, Facebook, or other social media platforms. Collected data must be labelled for clarity and manual annotation of huge number of text data is a difficult task. It is analyzed that many authors have used VEDAR and TextBlob technique for data labelling. The various data set used by existing authors are shown in Table 3. It is analyzed from Tables 3 that many authors employed only a subset of available data for their studies. So, resulting models may not be applicable to wide range of situations. The availability of verified COVID-19 datasets is a big challenge for the scientific community. Training and testing of model is performed by using small dataset due to unavailability of annotated dataset in early stages of research.

Table 4 shows the summarised COVID-19 and its vaccine-related research work of existing authors by using machine and deep learning models.

It can be analyzed from Table 4 that accuracy obtained by various machine learning model is between 75% and 99%. However, highest accuracy of 99.33% is achieved by using hybrid deep learning model of CNN and BiLSTM.

4.2. Motivations

Application of data analysis plays an important role in various fields. Analysis and evaluation of the people's sentiments towards serious diseases is a important application of natural language processing and data mining. Sentiment analysis is a simple, efficient, and effective way to evaluate public opinion on illnesses and their transmission.

4.3. Recommendations

The recommendations mostly indicate future tactics that can be implemented for advanced sentiment analysis research of various themes. The data analysis outcome depends on characteristics of data sets and applied techniques. Hindi and other Indian languages can also be used for sentiment analysis. Some additional recommendations that can be considered for advancement of sentiment analysis research:

- Multi-lingual Sentiment Analysis: multiple language based sentiment analysis can provide more comprehensive view of sentiments across different cultures and geographies. Therefore, development of sentiment analysis techniques for different Indian languages is a thrust area of research.
- Fine-grained Sentiment Analysis: Most existing sentiment analysis methods classify texts into positive, negative, and neutral. However, this approach oversimplifies the complexity of human emotions. Fine-grained sentiment analysis such as happiness, sadness, anger, and fear, can be considered.
- Domain-Specific Sentiment Analysis: Sentiment analysis techniques developed for one domain may not perform better on another domain due to different language, context, and cultural norms. Therefore, developing domain-specific sentiment analysis model can improve its accuracy and effectiveness for specific applications, such as customer reviews, political speeches, or social media conversations.
- Combination of Multiple Data Sources: Data from multiple sources such as social media, news articles, and survey responses can provide more comprehensive and diverse data set for sentiment analysis. Integrating these data sources can lead to more accurate and reliable sentiment analysis results.
- Integration of Deep Learning: Deep learning techniques have shown promising results of sentiment analysis. Therefore, integrating deep learning techniques with traditional machine-learning approaches can further improve the accuracy and effectiveness of sentiment analysis.

4.4. Future implications of research about COVID

Several potential outcomes are outlined below:

- Massive amounts of data can be used to investigate various systems.
- Sentiment analysis can be applied by using multiple languages.
- Income level and demographics can also be considered while analysing public's sentiments.
- Sentiment analysis can be performed according to different age groups of people.
- Collection of accurate spatial data is crucial for effective Geo-referencing.
- During crisis, more targeted analysis can be performed to support policymakers, governments, and communities.

5. Conclusions

Systematic literature review of the sentiment analysis of COVID-19 and its vaccination in INDIA during past three years is presented in this work. The applications of lexicons, machine, and deep Learning techniques for sentiment analysis of COVID-19 and its vaccination are analyzed. Various data sources, available data set and methods applied by the various authors are discussed. The common problems with collected data and limitations of the existing works are also presented. It is concluded that data set of multiple languages can be utilized for effective and accurate analysis. Hybrid model of deep and machine learning can be utilized to analyze massive data from various sources.

References

[1] S. Agarwal and S. Sarkar, Topical analysis of migration coverage during lockdown in india by mainstream print media, *Plos One* **17** (2022), e0263787.

[2] P. Aggrawal, B.L.K. Jolly, A. Gulati, A. Sethi, P. Kumaraguru and T. Sethi, Psychometric analysis and coupling of emotions between state bulletins and twitter in india during covid-19 infodemic, *Frontiers in Communication* **6** (2021), 695913.

[3] A.H. Alamoodi, B.B. Zaidan, A.A. Zaidan, O.S. Albahri, K. Mohammed, R.Q. Malik, E.M. Almahdi, M.A. Chyad, Z. Tareq, A.S. Albahri, et al., Sentiment analysis and its applications in fighting covid-19 and infectious diseases: A systematic review, Expert systems with applications 167 (2021), 114155.

[4] Alec Go and L.H. Richa Bhayani, A twitter sentiment analysis tool, 2020. URL: https://help.sentiment140.com/home

[5] W. Aljedaani, E. Saad, F. Rustam, I. de la Torre Diez and I. Ashraf, Role of artificial intelligence for analysis of covid-19 vaccination-related tweets: Opportunities, challenges, and future trends, Mathematics 10 (2022), 3199.

[6] R.R. Aryal and A. Bhattarai, Sentiment analysis on covid-19 vaccination tweets using naive bayes and lstm, Advances in Engineering and Technology: An International Journal 1 (2021), 57–70.

[7] S. Atef and A.B. Eltawil, Assessment of stacked unidirectional and bidirectional long short-term memory networks for electricity load forecasting, Electric Power Systems Research 187 (2020), 106489.

[8] G. Barkur, G.B. Kamath, et al., Sentiment analysis of nationwide lockdown due to covid 19 outbreak: Evidence from india, *Asian journal of psychiatry* **51** (2020), 102089.

[9] A. Borah, Detecting covid-19 vaccine hesitancy in india: a multimodal transformer based approach, *Journal of Intelligent Information Systems* (2022), 1–17.

[10] L. Breiman, Random forests, *Machine Learning* **45** (2001), 5–32.

[11] L. Cao and Q. Liu, Covid-19 modeling: A review, *medRxiv*, (2022).

[12] Y. Cao, M. Long, J. Wang, Q. Yang and P.S. Yu, Deep visual-semantic hashing for cross-modal retrieval, in: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (2016), pp. 1445–1454.

[13] S. Chahar and P.K. Roy, Covid-19: A comprehensive review of learning models, *Archives of Computational Methods in Engineering* (2021), 1–26.

[14] K. Chakraborty, S. Bhatia, S. Bhattacharyya, J. Platos, R. Bag and A.E. Hassanien, Sentiment analysis of covid-19 tweets by deep learning classifiers—a study to show how popularity is affecting accuracy in social media, *Applied Soft Computing* **97** (2020), 106754.

[15] D. Chehal, P. Gupta and P. Gulati, Covid-19 pandemic lockdown: An emotional health perspective of indians on twitter, *International Journal of Social Psychiatry* **67** (2021), 64–72.

[16] N. Chintalapudi, G. Battineni and F. Amenta, Sentimental analysis of covid-19 tweets using deep learning models, *Infectious Disease Reports* **13** (2021), 329–339.

[17] M. Chiny, M. Chihab, O. Bencharef and Y. Chihab, Lstm, vader and tf-idf based hybrid sentiment analysis model, *International Journal of Advanced Computer Science and Applications* **12** (2021).

[18] J. Cruz-Cardenas, E. Zabelina, J. Guadalupe-Lanas, A. Palacio-Fierro and C. Ramos-Galarza, Covid-19, consumer behavior, technology, and society: A literature review and bibliometric analysis, *Technological Forecasting and Social Change* **173** (2021), 121179. URL: https://www.sciencedirect.com/science/article/pii/S0040162521006120, doi: https://doi.org/10.1016/j.techfore.2021.121179

[19] A.D. Dubey, Public sentiment analysis of covid-19 vaccination drive in india, Available at: *SSRN 3772401*. (2021).

[20] E. Ermatita, A. Abdiansah, D.P. Rini and F. Febry, Sentiment analysis of covid-19 using multimodal fusion neural networks, *TEM Journal* **11** (2022), 1316–1321.

[21] P. Ghasiya and K. Okamura, Investigating covid-19 news across four nations: A topic modeling and sentiment analysis approach, *IEEE Access* **9** (2021), 36645–36656.

[22] S. Ghosh, S. Shankar, K. Chatterjee, K. Chatterjee, A.K. Yadav, K. Pandya, V. Suryam, S. Agrawal, S. Ray, V. Phutane, et al., Covishield (azd1222) vaccine effectiveness among healthcare and frontline workers of indian armed forces: Interim results of vin-win cohort study, *Medical Journal Armed Forces India* **77** (2021), S264–S270.

[23] K. Gulati, S.S. Kumar, R.S.K. Boddu, K. Sarvakar, D.K. Sharma and M. Nomani, Comparative analysis of Machine learning-based classification models using sentiment classification of tweets related to covid-19 pandemic, *Materials Today: Proceedings* **51** (2022), 38–41.

[24] P. Gupta, S. Kumar, R.R. Suman and V. Kumar, Sentiment analysis of lockdown in india during covid-19: A case study on twitter, *IEEE Transactions on Computational Social Systems* **8** (2020), 992–1002.

[25] V. Gupta, N. Jain, P. Katariya, A. Kumar, S. Mohan, A. Ahmadian and M. Ferrara, An emotion care model using multimodal textual analysis on covid-19, *Chaos, Solitons & Fractals* **144** (2021), 110708.

[26] K. Hall, V. Chang and C. Jayne, A review on natural language processing models for covid-19 research, *Healthcare Analytics* (2022), 100078.

[27] A.S. Imran, S.M. Daudpota, Z. Kastrati and R. Batra, Cross-cultural polarity and emotion detection using sentiment analysis and deep learning on covid-19 related tweets, *IEEE Access* **8** (2020), 181074–181090.

[28] W. India, Who india weekly covid-19 situational report, edition 116 published july 22, (2022). URL: https://cdn.who.int/media/docs/default-source/wrindia/situation-report/india-situation-report-116.pdf?sfvrsn=1a5f2a59_2

[29] H. Jelodar, Y. Wang, R. Orji and S. Huang, Deep sentiment classification and topic discovery on novel coronavirus or covid-19 online discussions: Nlp using lstm recurrent neural network approach, *IEEE Journal of Biomedical and Health Informatics* **24** (2020), 2733–2742.

[30] H. Kaur, S.U. Ahsaan, B. Alankar and V. Chang, A proposed sentiment analysis deep learning algorithm for analyzing covid-19 tweets, *Information Systems Frontiers* (2021), 1–13.

[31] M. Khan and A. Malviya, Big data approach for sentiment analysis of twitter data using hadoop framework and deep learning, in: *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), IEEE*, (2020), pp. 1–5.

[32] V. Krkova, Y. Manolopoulos, B. Hammer, L. Iliadis and I. Maglogiannis, Artificial Neural Networks and *Machine learning*–ICANN 2018: *27th International Conference on Artificial Neural Networks*, Rhodes, Greece, October 4–7, 2018, Proceedings, Part III. volume **11141**. Springer, (2018).

[33] A. Kumar, J.P. Singh, N.P. Rana and Y.K. Dwivedi, Multi-channel convolutional neural network for the identification of eyewitness tweets of disaster, *Information Systems Frontiers* (2022), 1–16.

[34] V. Kumar, Spatiotemporal sentiment variation analysis of geotagged covid-19 tweets from india using a hybrid deep learning model, *Scientific Reports* **12** (2022), 1–14.

[35] A. Kumari, P. Ranjan, S. Chopra, D. Kaur, T. Kaur, K.B. Kalanidhi, A. Goel, A. Singh, U. Baitha, B. Prakash and N.K. Vikram, What indians think of the covid-19 vaccine: A qualitative study comprising focus group discussions and thematic analysis, *Diabetes & Metabolic Syndrome: Clinical Research & Reviews* **15** (2021), 679–682. URL: https://www.sciencedirect.com/science/article/pii/S1871402121000953, doi: https://doi.org/10.1016/j.dsx.2021.03.021

[36] R. Lamsal, Design and analysis of a large-scale covid-19 tweets dataset, *Applied Intelligence* **51** (2021), 2790–2804.

[37] J. Li and E. Hovy, Reflections on sentiment/opinion analysis, in: *A practical guide to sentiment analysis*, Springer, (2017), pp. 41–59.

[38] D.V. Lindley, Fiducial distributions and bayes' theorem, *Journal of the Royal Statistical Society, Series B (Methodological)* (1958), 102–107.

[39] S. Majumder, A. Aich and S. Das, Sentiment analysis of people during lockdown period of covid-19 using svm and logistic regression analysis. Available at: *SSRN 3801039*. (2021).

[40] K. McKeown, A. Agarwal and F. Biadsy, Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams, (2009).

[41] S. Mendon, P. Dutta, A. Behl and S. Lessmann, A hybrid approach of *Machine learning* and lexicons to sentiment analysis: enhanced insights from twitter data of natural disasters, *Information Systems Frontiers* **23** (2021), 1145–1168.

[42] T.T. Mengistie and D. Kumar, Deep learning based sentiment analysis on covid-19 public reviews, in: *2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), IEEE*, (2021), pp. 444–449.

[43] A.A. Mir and R. Sevukan, Sentiment analysis of indian tweets about covid-19 vaccines, *Journal of Information Science* 01655515221118049, (2022).

[44] S. Mishra, A. Verma, K. Meena and R. Kaushal, Public reactions towards covid-19 vaccination through twitter before and after second wave in india, *Social Network Analysis and Mining* **12** (2022), 1–16.

[45] P. Misra and J. Gupta, Impact of covid 19 on indian migrant workers: decoding twitter data by text mining, *The Indian Journal of Labour Economics* **64** (2021), 731–747.

[46] S. Patel, A comprehensive analysis of convolutional neural network models, *International Journal of Advanced Science and Technology* **29** (2020), 771–777.

[47] T.R. Patil, Msss performance analysis of naive bayes and j48 classification algorithm for data classification, *Intl Journal of Computer Science and Applications* **6** (2013).

[48] A.N. Prabhu, G.B. Kamath, D.V. Pai, et al., Keeping the country positive during the covid 19 pandemic: Evidence from india, *Asian Journal of Psychiatry* **51** (2020), 102118.

[49] S. Praveen, R. Ittamalla and G. Deepak, Analyzing the attitude of indian citizens towards covid-19 vaccine–a text analytics study, *Diabetes & Metabolic Syndrome: Clinical Research & Reviews* **15** (2021), 595–599.

[50] G. Preda, covid-19-tweets data-set, (2020). URL: https://github.com/gabrielpreda/covid-19-tweets

[51] Q. Qiu, Z. Xie, L. Wu and L. Tao, Dictionary-based automated information extraction from geological documents using a deep learning algorithm, *Earth and Space Science* **7** (2020), e2019EA000993.

[52] K. Roy, S. Kar and R.N. Das, Chapter 6 – selected statistical methods in qsar, in: K. Roy, S. Kar, Das, R.N. (Eds.), *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*, Academic Press, Boston, (2015), pp. 191–229. URL: https://www.sciencedirect.com/science/article/pii/B9780128015056000065, doi: https://doi.org/10.1016/B978-0-12-801505-6.00006-5

[53] F. Rustam, M. Khalid, W. Aslam, V. Rupapara, A. Mehmood and G.S. Choi, A performance comparison of supervised Machine learning models for covid-19 tweets sentiment analysis, *Plos One* **16** (2021), e0245909.

[54] E. Saad, S. Din, R. Jamil, F. Rustam, A. Mehmood, I. Ashraf and G.S. Choi, Determining the efficiency of drugs under special conditions from users' reviews on healthcare web forums, *IEEE Access* **9** (2021), 85721–85737.

[55] R.M. Sallam, M. Hussein and H.M. Mousa, Improving collaborative filtering using lexicon-based sentiment analysis, *International Journal of Electrical and Computer Engineering* **12** (2022), 1744.

[56] J. Samuel, G.M.N. Ali, M.M. Rahman, E. Esawi and Y. Samuel, Covid-19 public sentiment insights and Machine learning for tweets classification, *Information* **11** (2020), 314.

[57] C. Singh, T. Imam, S. Wibowo and S. Grandhi, A deep learning approach for sentiment analysis of covid-19 reviews, *Applied Sciences* **12** (2022a), 3709.

[58] M. Singh, H.K. Dhillon, P. Ichhpujani, S. Iyengar and R. Kaur, Twitter sentiment analysis for covid-19 associated mucormycosis, *Indian Journal of Ophthalmology* **70** (2022b), 1773–1779.

[59] M. Singh, A.K. Jakhar and S. Pandey, Sentiment analysis on the impact of coronavirus in social life using the bert model, *Social Network Analysis and Mining* **11** (2021), 1–11.

[60] D. Sunitha, R.K. Patra, N. Babu, A. Suresh and S.C. Gupta, Twitter sentiment analysis using ensemble based deep learning model towards covid-19 in india and european countries, *Pattern Recognition Letters* **158** (2022), 164–170.

[61] P. Sv, J. Tandon, H. Hinduja, et al., Indian citizen's perspective about side effects of covid-19 vaccine–a Machine learning study, *Diabetes* & *Metabolic Syndrome: Clinical Research* & *Reviews* **15** (2021), 102172.

[62] A.S.M. Venigalla, S. Chimalakonda and D. Vagavolu, Mood of india during covid-19-an interactive web portal based on emotion analysis of twitter data, in: *Conference companion publication of the 2020 on computer supported cooperative work and social computing*, pp. 65–68. (2020).

[63] M. Wang, J. Xie, Z. Tan, J. Su, D. Xiong and L. Li, Towards linear time neural machine translation with capsule networks, *arXiv preprint arXiv:1811.00287*, (2018).

[64] S. Wang, C. Shao, J. Zhang, Y. Zheng and M. Meng, Traffic flow prediction using bi-directional gated recurrent unit method, *Urban Informatics* **1** (2022), 16.

[65] T. Wang, K. Lu, K.P. Chow and Q. Zhu, Covid-19 sensing: negative sentiment analysis on social media in china via bert model, *IEEE Access* **8** (2020), 138162–138169.

[66] WHO, Covid-19 weekly epidemiological update,edition 104 published 10 august, (2022). URL: https://help.sentiment140.com/home

[67] R. Xie, S.K.W. Chu, D.K.W. Chiu and Y. Wang, Exploring public response to covid-19 on weibo with lda topic modeling and sentiment analysis, *Data and Information Management* **5** (2021), 86–99.

[68] G. You, S. Gan, H. Guo and A.A. Dagestani, Public opinion spread and guidance strategy under covid-19: A sis model analysis, *Axioms* **11** (2022), 296.

[69] X. Zhu, M. Zhang, Y. Hong and R. He, Natural Language Processing and Chinese Computing: 9th CCF International Conference, NLPCC 2020, Zhengzhou, China, October 14–18, 2020, Proceedings, Part I. volume **12430**. Springer Nature, (2020).