Article

Evaluating Polarity Trend Amidst the Coronavirus Crisis in Peoples' Attitudes toward the Vaccination Drive

Rakhi Batra [1], Ali Shariq Imran [2,*], Zenun Kastrati [3], Abdul Ghafoor [1], Sher Muhammad Daudpota [1] and Sarang Shaikh [1]

[1] Department of Computer Science, Sukkur IBA University, Sukkur 65200, Pakistan; rakhi@iba-suk.edu.pk (R.B.); aghafoor.mscsf19@iba-suk.edu.pk (A.G.); sher@iba-suk.edu.pk (S.M.D.); sarang.msse17@iba-suk.edu.pk (S.S.)

[2] Department of Computer Science (IDI), Norwegian University of Science & Technology (NTNU), 2815 Gjøvik, Norway

[3] Department of Informatics, Linnaeus University, 351 95 Växjö, Sweden; zenun.kastrati@lnu.se

* Correspondence: ali.imran@ntnu.no

Abstract: It has been more than a year since the coronavirus (COVID-19) engulfed the whole world, disturbing the daily routine, bringing down the economies, and killing two million people across the globe at the time of writing. The pandemic brought the world together to a joint effort to find a cure and work toward developing a vaccine. Much to the anticipation, the first batch of vaccines started rolling out by the end of 2020, and many countries began the vaccination drive early on while others still waiting in anticipation for a successful trial. Social media, meanwhile, was bombarded with all sorts of both positive and negative stories of the development and the evolving coronavirus situation. Many people were looking forward to the vaccines, while others were cautious about the side-effects and the conspiracy theories resulting in mixed emotions. This study explores users' tweets concerning the COVID-19 vaccine and the sentiments expressed on Twitter. It tries to evaluate the polarity trend and a shift since the start of the coronavirus to the vaccination drive across six countries. The findings suggest that people of neighboring countries have shown quite a similar attitude regarding the vaccination in contrast to their different reactions to the coronavirus outbreak.

Keywords: coronavirus; COVID-19; pandemic; polarity assessment; opinion mining; emotion detection; Twitter posts; BERT; GloVe; DNN; LSTM; FastText; global crisis

1. Introduction

COVID-19 is an infectious disease and the first case was reported in December 2019 in Wuhan city of China. It rapidly spread around the globe and declared as pandemic on 11 March 2020 by World Health Organization (WHO) (https://www.who.int/news/item/27-04-2020-who-timeline---covid-19, accessed on 23 March 2021). As of 22 March 2021, the pandemic infected 123,868,982 people and 2,727,738 deaths have been reported around the globe according to Worldometer (https://www.worldometers.info/coronavirus/, accessed on 23 March 2021). The USA, Brazil, and India are the worst affected countries in terms of both case count and mortality (https://www.nationalgeographic.com/science/graphics/mapping-coronavirus-infections-across-the-globe, accessed on 25 March 2021) as shown in the Figure 1. Multiple variants of the coronavirus has been detected for instances UK and South African variants. On 14 December 2020, UK authorities notified the WHO about the coronavirus variant and initial studies investigated this variant may spread rapidly people to people. Researchers have stated that the COVID-19 variant first time reported in the UK is up to 100 percent more fatal than earlier strains (https://www.aljazeera.com/news/2021/3/10/uk-covid-19-variant-30-100-more-deadly-study-finds, accessed on 26 March 2021).

	Cases	Deaths
United States	29,692,901	539,207
Brazil	11,780,820	287,499
India	11,514,331	159,370
Russia	4,378,656	92,266
United Kingdom	4,287,463	126,035
France	4,152,604	91,148
Italy	3,306,711	103,855
Spain	3,212,332	72,910
Turkey	2,950,603	29,777
Germany	2,639,258	74,402
Canada	928,283	22,577
Sweden	738,537	13,236
Pakistan	619,259	13,757
Norway	84,553	648

Figure 1. COVID-19 cases and deaths reported country-wise, 19 March 2021.

COVID-19 emergencies have affected the individual mental health causing insecurity, emotional isolation, confusion, and depression due to loss in business, education, and work [1]. This pandemic situation changed the normal routine of people around the world such as academic activities shifted from physical to online mode, change in the way people interact daily, conduct business or do shopping. Although it disturbed all the activities, people from different cultures did not react and respond to the pandemic in the same way. Our previous study has discussed this cultural difference concerning COVID-19 outbreak [2]. Twitter data of six countries from three different continents were collected to explore the emotions of people from different cultures about the decisions their respective governments took to control the coronavirus outbreak. The selected countries were India and Pakistan from Asia, Sweden and Norway from Europe, and USA and Canada from North America. Experimental results showed a high correlation between emotions from India and Pakistan and USA and Canada. Whereas, Norway and Sweden being neighboring countries with many cultural similarities showed the opposite polarity trends.

Almost after one year, now many countries worldwide have rolled out the COVID-19 vaccine to cure this infectious disease. Western countries are leading in the COVID-19 vaccination whereas African countries are lagging as can be depicted from Figure 2. United Kingdom (UK) became the first nation in the world to approve the BioNTech-Pfizer vaccine and a UK Grandmother Margaret Keenan has became the first person in the world to receive COVID-19 vaccine on 8 December 2020. Both the USA and Canada have started mass COVID-19 vaccination program outside a clinical trial on 14 December 2020. Sandra Lindsay was the first American vaccinated at Long Island Jewish Medical center and the first person from Canada was Anita Quidangen, a personal support worker injected in Toronto. Nordic neighboring countries Sweden and Norway rolled out the coronavirus vaccine drive on 27 December 2020. A 67-year-old Svein Andersen was the first person in Norway to receive the vaccine and from Sweden Gunn-Britt Johnsson, the 91-year-old woman was the first person. Manish Kumar, a hospital cleaning worker was the first Indian to receive vaccine on 16 January 2021. Pakistan kicks off vaccination on 2 February 2021, and Rana Imran Sikander from PIMS hospital Islamabad was the first person to receive the vaccine.
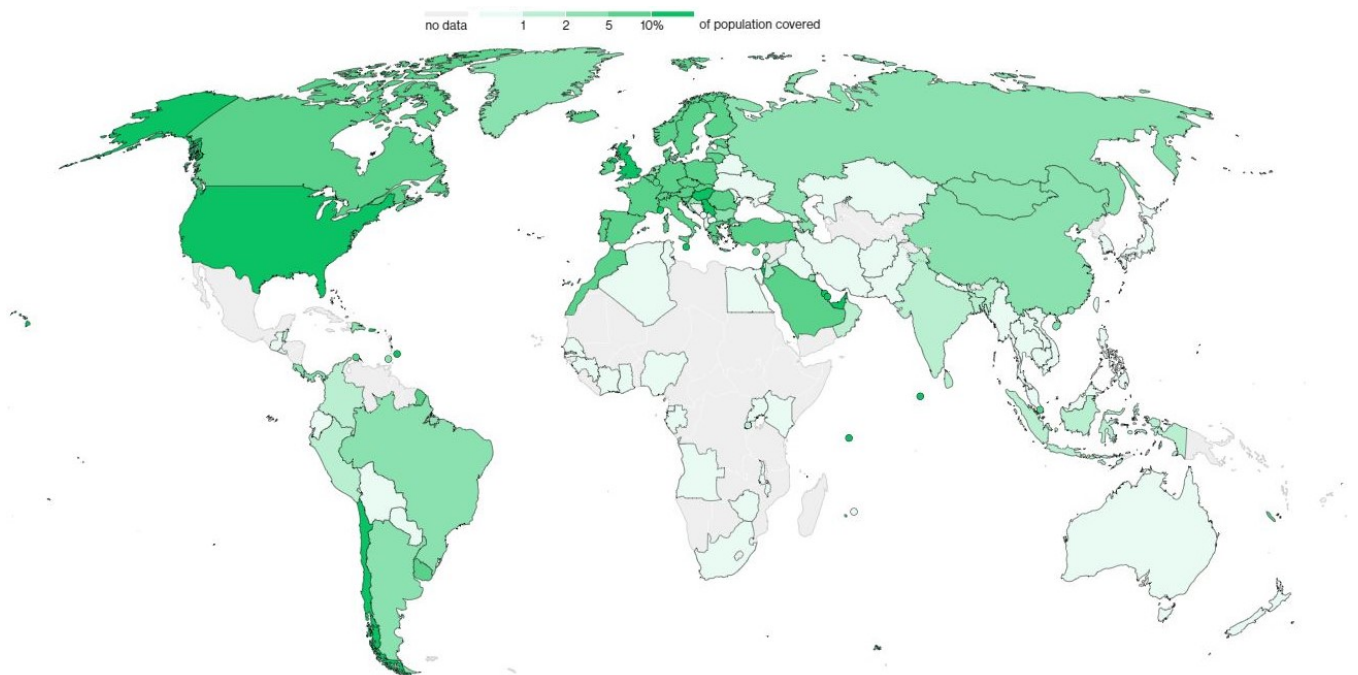
Figure 2. COVID-19 vaccine doses covered population, 24 March 2021 (Source Bloomberg).

According to Bloomberg vaccine tracker (https://www.bloomberg.com/graphics/covid-vaccine-tracker-global-distribution/, accessed on 25 March 2021) as of 24 March 2021, more than 468 million COVID-19 shots have been given across 135 countries. USA is leading with more than 128 million doses which cover the 19.7% of USA population. Canada has vaccinated 4.2 million people. India has vaccinated 50 million people and its bordering country Pakistan has injected just 325,000 doses. Sweden has vaccinated 1.4 million people and Norway has vaccinated 771,000 people. Few countries have also reported the side effects of COVID-19 vaccine. On 18 February 2021, Norwegian Medicine Agency acknowledged more than 1200 side effect reports (https://tinyurl.com/db2x86j7, accessed on 15 March 2021). Two Swedish regions (https://tinyurl.com/2empedan, accessed on 15 March 2021) stopped vaccination after receiving side effects reports on 14 February 2021. In earlier March, following Denmark, including Norway and other Nordic and central European countries halted giving AstraZeneca vaccines shots to its citizen amid deaths due blood clotting as a side-effect.

People generally are quick in sharing such news and personal experiences over social networks, and to base their opinions upon what they hear. Many would react and express various sentiments while commenting. The paper is motivated by the fact that such trends could pick up quickly—social trends could easily turn into mass gatherings and protests which ultimately turn into chaos as was observed in Arab spring. Timely analysis of people's sentiment on social platforms could help avoid such a situation and sentiment analysis is an efficient tool to automatically examine sentiment expressed in social media. Deep neural networks, especially LSTM networks and its different variants have shown good promise to process text for sentiment polarity extraction. The performance of the task has also benefited hugely from pretrained word embedding like GloVe, FastText, BERT, etc. Our previous study [2] has demonstrated the potential of these networks to extract sentiments related to COVID-19 from tweets posted from six countries, i.e., Pakistan, India, Norway, Sweden, the USA and Canada. The purpose of this study, therefore, is to detect changes in polarity and emotions of people after the launch of vaccine and its side effects expressed in tweets, and to find connection between the events that took place during the vaccination drive across various countries and emotions expressed on social networks. We proposed to utilize deep natural language models to analyse the tweets for sentiment polarity as well as emotion detection.

The key contributions of this study are:

1. Collection of tweets on COVID-19 related hashtags for the period of two months during the vaccination drive to analyze sentiment polarity and emotions.
2. Providing insights into the collective reactions amidst second wave, and to establish links with on-going events.
3. Finding correlation between emotions expressed at the start of the COVID-19 and the vaccination drive after a year for six countries across three continents.
4. Analysing polarity and emotions via state-of-the-art deep learning based NLP models trained on benchmark data sets Sentiment140 for polarity assessment and Emotion-Tweet for emotion classification and tested the model on COVID-19 Tweets.

The rest of the paper is organized in following manner. Section 2 presents the related work. Methodology Section 3 describes the model used to study people's attitudes from their tweets posted on Twitter. Results and their analysis are presented in Section 4, whereas the conclusion is drawn in Section 5.

2. Related Work

A recent development in sentiment analysis and affective computing is to explore textual data to get public views on financial markets [3], politics [4], education [5,6], etc., just to name a few. Various research studies have also discussed the people's reactions to events expressed in social media, in general, and Twitter in particular. Types of events include pandemic [7], protest [8], criminal and terrorist events [9], natural disasters [10], healthcare-related events [11], and so forth [12,13].

Many research studies have been conducted for different reasons including investigation of Twitter data to find the spreading pattern information on Ebola [14] and on the COVID-19 outbreak [15], track and know the public views on Twitter amid pandemic [16,17], examine the intuitions that Global Health can draw from social networks [18], and the reaction of people from different nations during the pandemic, toward the actions their respective governments took to control the coronavirus outbreak [2]. Fung et al. [19] investigated people's reactions toward the Ebola outbreak on Twitter and Google. Experimental results showed a majority of emotions express the negative sentiment. The authors in [20] examine people's emotional answers during the Middle East Respiratory Syndrome (MERS) outbreak in South Korea. They found that 80% of tweets were neutral. Anger increased over time. The majority of people were blaming the Korean government and a decline in fear and sadness tweets were reported over time.

Many sentiment analysis studies related to COVID-19 have been done based on the social media data as shown in Figure 3, mainly focused on sentiment analysis concerning the use of masks [21], fake information detection [22], emotion classification [23], polarity detection [24], depression monitoring [25], Tourism [26] and so on.

2.1. Sentiment Polarity Assessment on COVID-19 Data

Research has been done to classify the sentiment polarity of Twitter data for the coronavirus. Sakun et al. in research paper [27] have explored the Twitter trends related to COVID-19. They collected 107,990 English tweets about the coronavirus and used sentiment analysis and topic modeling to explore the tweets. Experiment results showed three main aspects of tweets. (1) trends related to symptoms and the spread of COVID-19 can be divided into three stages. (2) Sentiment analysis reveals that most people's views were negative about Coronavirus. (3) COVID-19 tweets were divided into three topics namely: the COVID-19 pandemic emergency, how to control COVID-19, and reports on COVID-19. Barkur et al. [28] explored the Twitter data for sentiments of people in India about COVID-19 lockdown, and observation showed that the majority of views about lockdown were negative but also there were some positive opinions. In another research study [29], the authors have proposed the machine learning model to predict an individual's awareness of the protective measures against the coronavirus in Saudi Arabia. In this study, Arabic tweets related to COVID-19 were collected and machine learning

models: support vector machine, *K*-nearest neighbors, and naïve Bayes were used to train and test the Arabic tweets, SVM model outperformed with an accuracy of 85%.
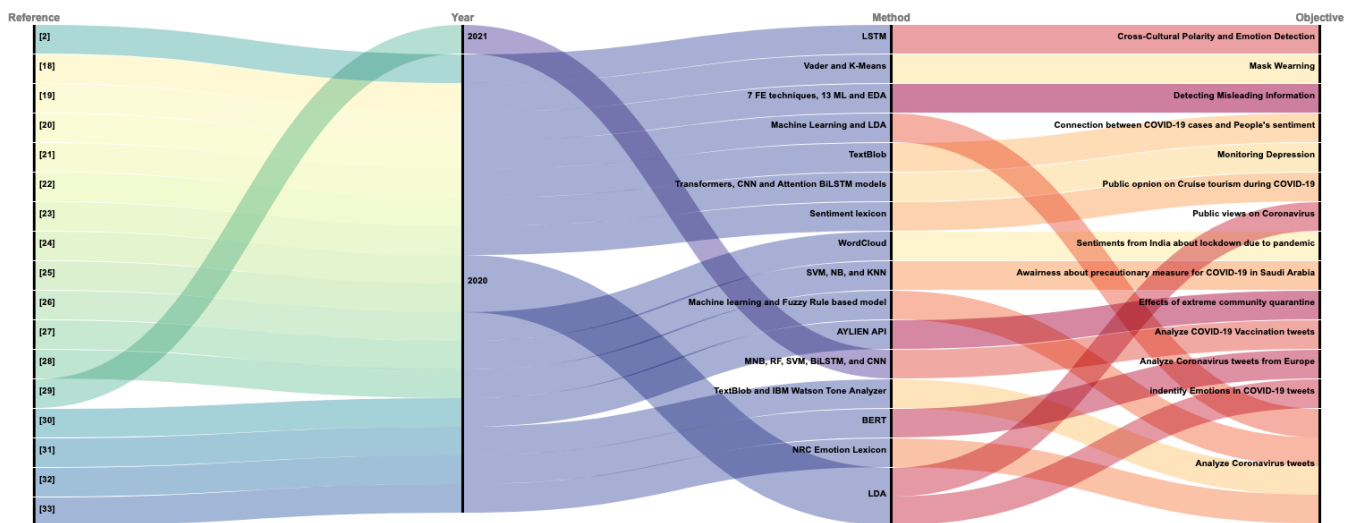
Figure 3. Comparison of sentiment analysis studies related to COVID-19 on Twitter data.

The research article [30] has proposed the deep learning model for sentiment analysis of coronavirus tweets. The study has collected two types of tweets: (1) 23,000 most retweeted tweet collected between 1 January 2020 to 23 March 2020, tweets were explored and results reveal that the maximum number of the tweets were neutral and negative and (2) 226,668 tweets gathered between December 2019 and May 2020 show the maximum number of tweets were positive and neutral tweets. The study concluded overall reaction of people about COVID-19 on Twitter was positive yet citizens retweeted mostly negative tweets. The authors in the paper [24] have investigated the relationship between the sentiment of public and coronavirus cases. The study used the TextBlob sentiment corpus to compute the polarity of tweets. Results reveal that there is a connection between the sentiment of the public and COVID-19 cases. Important events such as government regulation to slowdown spread, a celebration of important days can affect the people's sentiment. The study showed a weak correlation between sentiment polarity and that increase in numbers of COVID-19 cases, public sentiment is affected but not that much by the increase of coronavirus cases.

Pastor et al. in paper [31] have explored the Twitter sentiment analysis to classify the views of Filipinos on extreme community quarantine measures announced by the Philippines government to slow the spread of coronavirus. Sentiment results revealed that food supply and support from government was major problem face by the people and it concluded that most of the people showed negative sentiment while some users also posted positive opinions. The authors of another research paper [32] analyzed people's reactions regarding the coronavirus vaccine. The study collected 2,349,659 tweets for a month once the first dose vaccinated in the UK. Experiment results point out that most of the tweets were neutral while tweets in favor of the vaccine overtook the tweets against the vaccine. Kaur et al. in their research paper [33] have collected 16,138 tweets from three different months of 2020 namely February, May, and June to monitor the polarity of tweets amid COVID-19. The number of negative tweets surpassed the neutral and positive tweets in all different time intervals as expected. Comparing the share of polarity classes from February to June, the negative tweets were decreased from 43.90% to 38.05% while the ratio of positive tweets increased from 21.38% to 27.01%. The share of the neutral tweets has nearly remained the same, 34.07% and 34.94%. The research study [34] has explored tweets from Europe regarding COVID-19. The authors collected 4.6 million geotagged

tweets from December 2019 to April 2020. Experimental results stated that as time passes a downward trend of the negative sentiment was observed.

2.2. Emotion Classification on COVID-19 Data

The authors in the study [2] have investigated the Twitter data of six countries from three different continents to know the emotions of people from different cultures about actions their respective governments have taken on COVID-19. Countries include India and Pakistan from Asia, Sweden and Norway from Europe, and the USA and Canada from North America. Deep Learning-based LSTM models are used to train and test data. The study reveals a high correlation in a tweet from India and Pakistan, and the USA and Canada. Although two Nordic countries have many cultural similarities, Norway and Sweden showed opposite emotions about COVID-19. The research study [35] has collected the tweets from twelve countries related to the coronavirus and explored the tweets to know people's opinions from different countries about COVID-19. Experimental results conclude that majority of people showed positive and hopeful thoughts but also fear, sadness, and disgust opinions were observed. However, the USA, France, the Netherlands, and Switzerland showed distrust and anger more than the other eight countries. Xue et al. [36] have analyzed the 11 sentiment analysis topic identified from 1.5 million tweets collected related to the coronavirus. The authors proposed a Latent Dirichlet Allocation (LDA) topic modeling algorithm to explore all topics. Experimental results found that fear is the dominant emotion in all topics.

3. Methodology

This section starts with explanation of our process of collecting tweets related to COVID-19 during the second wave of the coronavirus. We also elaborate the process of sentiment and emotion analysis on tweets from six countries including Pakistan, India, Norway, Sweden, the USA and Canada.

3.1. Data Set—Tweets Related to Second COVID-19 Wave

The data set used in this study contains tweets from Twitter for cross-cultural emotion recognition during the second wave of the coronavirus. For reliable cross culture polarity measurement, six countries were selected from three continents; two from each that share similar culture. The selected countries were India and Pakistan from Asia, Norway and Sweden from Europe, and Canada and the USA from North America. These six countries were chosen in particular to compare the trend between the polarity expressed during the first wave reported in [2] with the second wave during the vaccination drive.

Data Collection: Twitter provides API to extract bulk data from their platform for analysis. There are two types of API, i.e., Stream API and Search API. Stream API is used to get live data, whereas Search API is used to extract historical data (up to the last 7 days) by applying some filters. We used Twitter Search API known as Tweepy for collecting the required data set. As we aimed to analyze the peoples' sentiment over the progress of COVID-19 vaccine and second wave, we collected the data for a time period $T_p = S_d, E_d$, where S_d is start date of second wave and E_d is the end date. The following query was used to extract the data:

[Keyword] lang:[en/ur] until:E_d since:S_d -filter:links -filter:retweets

The keywords were selected such that they are directly linked to the coronavirus and seem to be trending on twitter since the start of virus. The keywords used for extracting tweets are: *lockdown*, *COVID19Pandemic*, *StayHomeSaveLives*, *stayhome*, *Covid_19*, *COVID*, *Coronavirus*, *secondwave*, *pandemic*, *covid19*, *vaccine*. Links and retweets were being filtered out to exclude the less informative and repetitive tweets. Extracted tweets were cataloged in an *xlsx* file as a raw data set, where each tweet record contains 72 fields

that describe tweet content and user information. For our objective we just retained six fields, i.e., *tweetid*, *tweettext*, *date*, *language*, *userid*, and *location*.

Data preparation: The raw data set was processed further to clean the tweet text up and to extract the emojis from it. In preprocessing, first we removed unnecessary symbols, spaces, and mentioned users from tweet text and then we used NLTK library to remove punctuation and stop-words and got the cleaned tweet text. As we aim to use this data set for emotion recognition, so to support the sentiment analyzer for accurate results we extracted the emojis from tweet text because emojis are true representation of users' reaction/emotions in any textual composition.

In the final dataset (https://tinyurl.com/u47h9y7t, accessed on 28 March 2021), each tweet was cataloged by *Tweetid*, *date*, *language*, *cleanedtext*, *emoji*, *sentimentscore*, *subjectivity*, *polarity*, *userid*, and *countrycode*. There are 801,692 tweets from six countries in the final data set. Country-wise distribution of tweets is shown in Table 1.

Table 1. Country-wise Distribution of Tweets.

Country	December-2020	January-2021	February-2021	Total
United States	131,254	317,016	177,950	626,220
Canada	12,171	60,389	39,456	112,016
India	18,772	21,350	11,862	51,984
Norway	489	2481	143	3113
Pakistan	1147	2627	1612	5386
Sweden	688	1332	953	2973
Total	164,521	405,195	231,976	801,692

3.2. Classification Models

As this work is an extension to our previous work [2], in order to assess change in peoples' sentiment and emotion after almost a year's time to our previous results, we keep the models same as our previous work. Readers are advised to consult section V in [2] for further details on algorithms for sentiment and emotion detection. Figure 4 shows the abstract model of the proposed classification system.

All three classifiers (A, B & C) are based on deep neural networks (DNN), Long Short-Term Memory (LSTM) Netowks and Convolution Neural Network (CNN).

Deep Neural Network (DNN): A DNN is a simplest form of neural networks. It's a layered architecture with all neurons at one layer fully connected with all neuron at next layer through an activation function.

Long Short Term Memory (LSTM) Network: Although fully connected deep neural networks are good at processing text and other small sequences, their performance degrades when sequences are longer. To address the issue of longer sequences, LSTM deep neural networks process current input and also retain previous state which is output from previous inputs. The capability of LSTM to retain previous state enables it to understand the word context; therefore, it is able to outperform DNN and other networks at processing long sequences.

Convolution Neural Network: A CNN deep neural network relies on two major operations, convolution and pooling. The convolution operation is performed on input text or image with filters of different sizes to produce feature map which can be further used for performing classification. The pooling operation involves sliding a two-dimensional filter over each channel of convoluted feature map to summarize features laying in sub-regions of the image or text. Traditionally, CNN is more appropriate for image processing, however recently it has also started showing enough promise on sequence processing too.

The *classifier A*, based on LSTM with pretrained FastText [37] embedding is trained on Sentiment140 [38] which contains a total number of 1.4 million tweets, equally distributed among positive and negative sentiment polarities. Table 2 shows the results of different models on Sentiment140 data set. The model based on LSTM and pretrained FastText

outperforms all other models. The summary of LSTM + FastText model is shown in Figure 5.
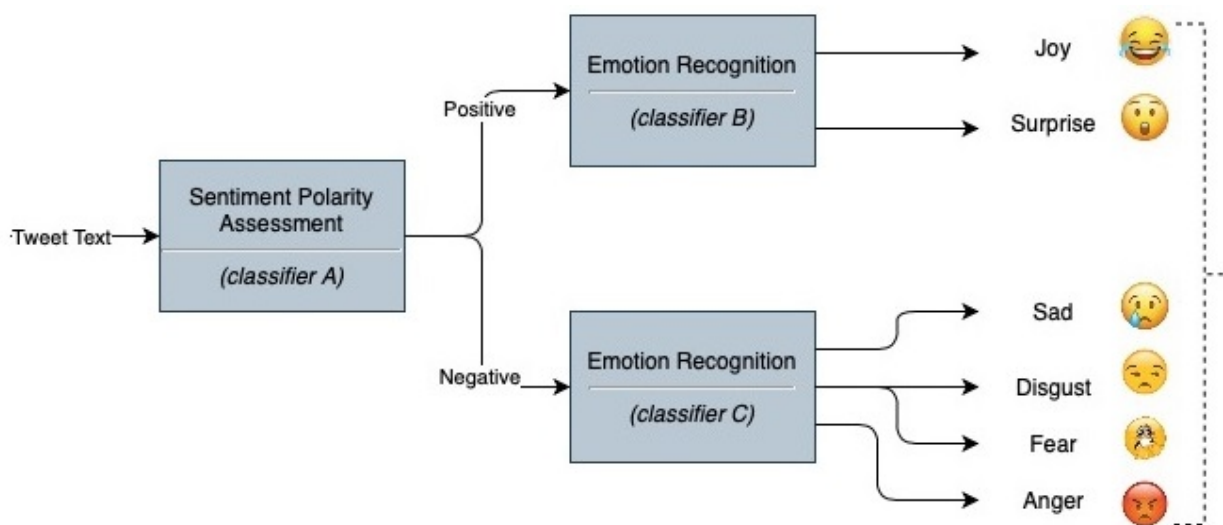

Figure 4. Abstract model for tweets' classification.

Table 2. F1 and accuracy scores of Six Deep Learning Models.

Model #	Model Name	F1 Score	Accuracy
1	DNN (Baseline)	79.0%	78.4%
2	**LSTM + FastText**	**82.4%**	**82.4%**
3	LSTM + GloVe	81.5%	81.4%
4	LSTM + GloVe Twitter	80.4%	80.4%
5	LSTM + w/o Pretrained Embed.	81.6%	81.4%
6	CONV Based on [39]	81.7%	81.1%

```
Layer (type)                 Output Shape              Param #
=================================================================
embedding_1 (Embedding)      (None, 280, 300)          12,000,000
_____
lstm_1 (LSTM)                (None, 32)                42,624
_____
dense_1 (Dense)              (None, 2)                 66
=================================================================
Total params: 12,042,690
Trainable params: 12,042,690
Non-trainable params: 0
_____
```

Figure 5. Summary of Classifier A for Sentiment Polarity Classification.

The positive polarity tweets are further checked for positive emotions (joy and surprise) through *classifier B*, whereas negative polarity tweets are forwarded to *classifier C* for negative emotions (sad, disgust, fear, anger). For both classifier B and C, six different models were assessed on an Emotional Tweet data set [40], and the summary of results for positive and negative emotions is shown in Tables 3 and 4, respectively. In both cases, LSTM with GloVe Twitter word embedding outperformed all other models; therefore, it is used for assessing tweets emotions. The summary of LSTM + GloVe Twitter model is shown in Figure 6.

Table 3. F1 and Accuracy Scores of Five Proposed Models on Positive Emotions (Joy and Surprise).

Model #	Model Name	F1 Score	Accuracy
1	DNN (Baseline)	62.7%	78.4%
2	LSTM + FastText	67.5%	80.8%
3	LSTM + GloVe	69.0%	80.3%
4	LSTM + GloVe Twitter	**69.9%**	**81.9%**
5	LSTM + w/o Pretrained Embed.	68.4%	79.8%

Table 4. F1 and Accuracy Scores of Five Models on Negative Emotions (Sad, Anger, Fear).

Model #	Model Name	F1 Score	Accuracy
1	DNN (Baseline)	59.0%	64.5%
2	LSTM + FastText	62.1%	66.0%
3	LSTM + GloVe	65.8%	67.7%
4	LSTM + GloVe Twitter	**69.2%**	**69.9%**
5	LSTM + w/o Pretrained Embed.	62.1%	66.0%

```
_____
Layer (type)                 Output Shape              Param #
=================================================================
embedding_1 (Embedding)      (None, 280, 300)          6,555,000

_____
lstm_1 (LSTM)                (None, 32)                42,624

_____
dense_1 (Dense)              (None, 2)                 66
=================================================================
Total params: 6,597,690
Trainable params: 6,597,690
Non-trainable params: 0
_____
```

Figure 6. Model Summary for Classifier B and C.

4. Results & Analysis

Figure 7 shows a side-by-side country-wise comparison of sentiment polarity detection for the investigated period of 2 months. The sentiments are normalized to the range of 0–1 by computing the sum of tweets per day over total number of tweets for a given country. As shown in graphs depicted in Figure 7, there were quite a few tweets concerning the vaccination posted over the second half of December 2020 and first half of January 2021. It can be noted that there were also only few days with no tweets. In particular, there were two days (i.e., 10 January and 24 January 2021) where no tweets have been posted for Norway and one day (i.e., 10 January 2021) for Sweden. It is also interesting to note that the number of tweets posted over this period of examination is rapidly increased only in the second half of January 2021, and this growing trend of tweets concerning vaccination drive is seen from the all six countries.

There is a sudden change in the emotions on particular days as shown Figure 7, especially on January 20 where the peak of both negative and positive emotions expressed in Twitter is registered. One possible reason for this could be the spread of new variant of the coronavirus. Multiple variants of the COVID-19 virus emerged at the end of 2020, most notably new variant first time detected in the UK (known as 20I/501Y.V1, VOC 202012/01, or B.1.1.7), and South Africa is (known as 20H/501Y.V2 or B.1.351) (https://www.cdc.gov/coronavirus/2019-ncov/more/science-and-research/scientific-brief-emerging-variants.html, accessed on 25 February 2021). These new variants quickly spread around the globe. Nordre Follo Municipality of Norway goes into lockdown after the British variant of the coronavirus spread on 22 January 2021. A new variant killed two nursing home residents and identified 22 employees at the Langhus center.
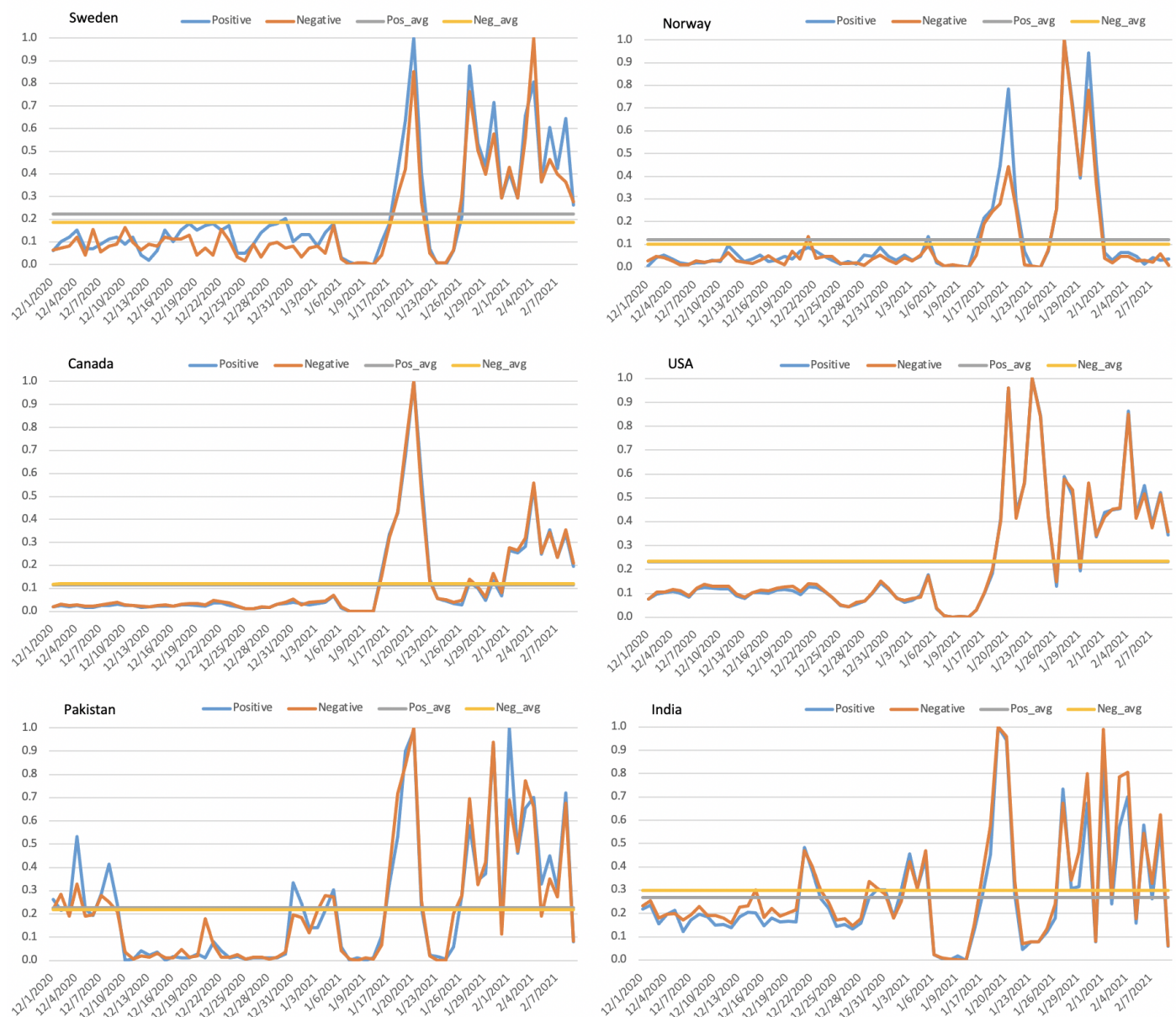
Figure 7. Side-by-side country-wise comparison of sentiments analysis on collected data for the period 1rd December 2020 to 9th February 2021. Cumulative positive and negative sentiment graphs along with the averaged tweets' polarity for Sweden (**top-left**), Norway (**top-right**), Canada (**middle-left**), USA (**middle-right**), Pakistan (**bottom-left**), and India (**bottom-right**).

Next, we analyzed the relationship between neighboring countries to see the sentiment polarity and emotion trend during the vaccination period. To achieve this, a Pearson's correlation between countries is computed, as shown in Table 5. The Pearson's correlation values indicate a high correlation in both positive and negative emotions of people from Pakistan and India (PK-IN), in contrast to people's sentiment toward vaccination drive in Canada and USA (US-CA), and Norway and Sweden (NO-SW). It is interesting to note that the Pearson's correlation between Norway and Sweden is 70% for positive and more than 60% for negative sentiments. This shows a higher correlation of sentiments about vaccination expressed in tweets on Twitter by the people of both countries, unlike their different sentiments about the coronavirus outbreak and lockdown reported in [2].

Table 5. Pearson's Correlation for Sentiment Polarity Between Neighbouring Countries.

No.	Correlation b/w	Positive	Negative
1	US-CA	0.623	0.624
2	PK-IN	0.837	0.865
3	NO-SW	0.703	0.616

Further, we examined the Pearson's correlation for emotions between neighbouring countries and a similar trend to sentiment polarity is observed. As can be seen in Table 6, the highest Pearson's correlation values across all the five emotions are shown for Pakistan and India, followed by the USA and Canada.

Table 6. Pearson's Correlation for Emotions Between Neighbouring Countries.

No	Correlation b/w	Joy	Surprise	Sad	Fear	Anger
1	US-CA	0.627	0.611	0.625	0.627	0.596
2	PK-IN	0.817	0.833	0.858	0.806	0.754
3	NO-SW	0.679	0.714	0.573	0.622	0.538

5. Conclusions and Future Work

This study aimed to analyze the emotions and sentiment polarity of people after the launch of vaccine and COVID-19 second wave. It also tried to show if there is any change in the sentiments of people since we studied the cross-cultural sentiment analysis in our previous study about one year ago. To achieve this objective, the same architecture was used from previous study which utilized the deep learning LSTM with pretrained embedding models to detect emotions from users' tweets on Twitter. Users' tweets were collected by querying the trending COVID-19 keywords from December 2020 to mid of February 2021 when different countries started to provide vaccine shots to public. In order to examine the change in sentiments of people from the start of virus, we limited the tweets from six countries that were used in previous study.

Result analysis showed that in December, people were mostly neutral about the vaccine and second wave but there was a sudden change in emotions after 15 January 2021. People started to express positive as well negative sentiments due to new variant of the coronavirus and governments' efforts toward the situation. We also applied Pearson's correlation to examine the emotion expression relationship between the neighbouring countries during the vaccination period. It indicated a high correlation in both positive and negative emotions of people from Pakistan and India (PK-IN), while people's sentiment toward vaccination drive in Canada and USA(US-CA) were 62% correlated, and in Norway and Sweden (NO-SW), the correlation was 70% for positive and 61% for negative despite of their different emotions during COVID-19 outbreak in 2020.

The study covered varying cultures including the EU, the USA, Canada and South Asian; however, it considered tweets only in English language. Usually, people in South Asia express their emotion using local languages like Urdu, Hindi, Sindh etc. The work can be extended in future to perform multilingual analysis for emotion and sentiment extraction from social media text related to COVID-19. Another trend which is popular on social media is the usage of roman Urdu, Hindi and other local languages. There is a strong need to consider this aspect of language when performing emotion and sentiment analysis for any topic of interest from social media.

Different transformer and attention based approaches for text processing have enormous potential to further improve accuracy of the proposed model. Usage of contextual word embedding like BERT, ELMo etc. are needed to be assessed for suitability in the task of social media text processing for sentiment and emotion analysis.

In this work, we have limited our focus on tweets, whereas other social media platforms like Facebook, Instagram etc. should be consider to learn more insights about people opinion related to COVID-19 and its vaccination process.

Finally, as they say, "a picture is worth a thousand word"; therefore, processing images for extracting people's sentiments and emotions could be considered another dimension of this work in future.

Author Contributions: R.B. prepared the data set by extracting and preprocessing tweets related to COVID-19. She also assisted Z.K. with the data analysis part. A.S.I. conceived the original idea, finalized contribution of this research, wrote the introduction part of paper and overall coordinate all the efforts of the research group on this paper. S.M.D. performed experiments on the tweets data set, wrote methodology part of the paper. S.S. contributed in literature review part and improved visualization and overall readability of the paper. Z.K. performed analysis on the results and led the whole results section along with multiple cycles of reviewing paper for improving its readability of manuscript. A.G. led the literature review part and performed multiple review of the paper to improve its readability. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data set used in this study can be found at https://tinyurl.com/u47h9y7t (accessed on 29 March 2021).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

COVID-19	Coronavirus disease 2019
PIMS	Pakistan Institute of Medical Sciences
WHO	World Health Organization
LSTM	Long short-term memory
GloVE	Global Vectors for Word Representation
BERT	Bidirectional encoder representations from transformers
DNN	Deep neural networks

References

1. Pfefferbaum, B.; North, C.S. Mental health and the Covid-19 pandemic. *N. Engl. J. Med.* **2020**, *383*, 510–512. [CrossRef] [PubMed]
2. Imran, A.S.; Daudpota, S.M.; Kastrati, Z.; Batra, R. Cross-cultural polarity and emotion detection using sentiment analysis and deep learning on COVID-19 related tweets. *IEEE Access* **2020**, *8*, 181074–181090. [CrossRef]
3. Carosia, A.; Coelho, G.P.; Silva, A. Analyzing the Brazilian financial market through Portuguese sentiment analysis in social media. *Appl. Artif. Intell.* **2020**, *34*, 1–19. [CrossRef]
4. Chauhan, P.; Sharma, N.; Sikka, G. The emergence of social media data and sentiment analysis in election prediction. *J. Ambient. Intell. Humaniz. Comput.* **2021**, *12*, 2601–2627. [CrossRef]
5. Kastrati, Z.; Imran, A.S.; Kurti, A. Weakly supervised framework for aspect-based sentiment analysis on students' reviews of MOOCs. *IEEE Access* **2020**, *8*, 106799–106810. [CrossRef]
6. Kastrati, Z.; Dalipi, F.; Imran, A.S.; Pireva Nuci, K.; Wani, M.A. Sentiment Analysis of Students' Feedback with NLP and Deep Learning: A Systematic Mapping Study. *Appl. Sci.* **2021**, *11*, 3986. [CrossRef]
7. Xiang, X.; Lu, X.; Halavanau, A.; Xue, J.; Sun, Y.; Lai, P.H.L.; Wu, Z. Modern senicide in the face of a pandemic: An examination of public discourse and sentiment about older adults and COVID-19 using machine learning. *J. Gerontol. Ser. B* **2021**, *76*, e190–e200. [CrossRef]
8. Won, D.; Steinert-Threlkeld, Z.C.; Joo, J. Protest activity detection and perceived violence estimation from social media images. In Proceedings of the 25th ACM International Conference on Multimedia, Mountain View, CA, USA, 23–27 October 2017; pp. 786–794.

9.   Burnap, P.; Williams, M.L.; Sloan, L.; Rana, O.; Housley, W.; Edwards, A.; Knight, V.; Procter, R.; Voss, A. Tweeting the terror: modelling the social media reaction to the Woolwich terrorist attack. *Soc. Netw. Anal. Min.* **2014**, *4*, 206. [CrossRef]

10.   Reynard, D.; Shirgaokar, M. Harnessing the power of machine learning: Can Twitter data be useful in guiding resource allocation decisions during a natural disaster? *Transp. Res. Part D Transp. Environ.* **2019**, *77*, 449–463. [CrossRef]

11.   Gohil, S.; Vuik, S.; Darzi, A. Sentiment analysis of health care tweets: review of the methods used. *JMIR Public Health Surveill.* **2018**, *4*, e43. [CrossRef] [PubMed]

12.   Dunkel, A.; Andrienko, G.; Andrienko, N.; Burghardt, D.; Hauthal, E.; Purves, R. A conceptual framework for studying collective reactions to events in location-based social media. *Int. J. Geogr. Inf. Sci.* **2019**, *33*, 780–804. [CrossRef]

13.   Kumar, A.; Jaiswal, A. Systematic literature review of sentiment analysis on Twitter using soft computing techniques. *Concurr. Comput. Pract. Exp.* **2020**, *32*, e5107. [CrossRef]

14.   Liang, H.; Fung, I.C.H.; Tse, Z.T.H.; Yin, J.; Chan, C.H.; Pechta, L.E.; Smith, B.J.; Marquez-Lameda, R.D.; Meltzer, M.I.; Lubell, K.M.; et al. How did Ebola information spread on twitter: broadcasting or viral spreading? *BMC Public Health* **2019**, *19*, 1–11. [CrossRef] [PubMed]

15.   Prabhakar Kaila, D.; Prasad, D.A. Informational flow on Twitter–Corona virus outbreak–topic modelling approach. *Int. J. Adv. Res. Eng. Technol. IJARET* **2020**, *11*, 128–134.

16.   Szomszor, M.; Kostkova, P.; St Louis, C. Twitter informatics: tracking and understanding public reaction during the 2009 swine flu pandemic. In Proceedings of the 2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, Lyon, France, 22–27 August 2021; IEEE: Piscataway, NJ, USA, 2011; Volume 1, pp. 320–323.

17.   Fu, K.W.; Liang, H.; Saroha, N.; Tse, Z.T.H.; Ip, P.; Fung, I.C.H. How people react to Zika virus outbreaks on Twitter? A computational content analysis. *Am. J. Infect. Control.* **2016**, *44*, 1700–1702. [CrossRef] [PubMed]

18.   Vorovchenko, T.; Ariana, P.; van Loggerenberg, F.; Amirian, P. # Ebola and Twitter. What insights can global health draw from social media? In *Big Data in Healthcare*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 85–98.

19.   Fung, I.C.H.; Tse, Z.T.H.; Cheung, C.N.; Miu, A.S.; Fu, K.W. Ebola and the social media. *Lancet* **2014**. [CrossRef]

20.   Do, H.J.; Lim, C.G.; Kim, Y.J.; Choi, H.J. Analyzing emotions in twitter during a crisis: A case study of the 2015 Middle East Respiratory Syndrome outbreak in Korea. In Proceedings of the 2016 International Conference on Big Data and Smart Computing (BigComp), Hong Kong, China, 18–20 January 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 415–418.

21.   Sanders, A.C.; White, R.C.; Severson, L.S.; Ma, R.; McQueen, R.; Paulo, H.C.A.; Zhang, Y.; Erickson, J.S.; Bennett, K.P. Unmasking the conversation on masks: Natural language processing for topical sentiment analysis of COVID-19 Twitter discourse. *medRxiv* **2021**. [CrossRef]

22.   Elhadad, M.K.; Li, K.F.; Gebali, F. COVID-19-FAKES: A Twitter (Arabic/English) dataset for detecting misleading information on COVID-19. In *International Conference on Intelligent Networking and Collaborative Systems*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 256–268.

23.   Xue, J.; Chen, J.; Hu, R.; Chen, C.; Zheng, C.; Su, Y.; Zhu, T. Twitter Discussions and Emotions About the COVID-19 Pandemic: Machine Learning Approach. *J. Med. Internet Res.* **2020**, *22*, e20550. [CrossRef]

24.   Luu, T.J.P.; Follmann, R. The Relationship between Sentiment Score and COVID-19 Cases in the USA 2020. Available online: https://jackluu.io/files/LuuResearchPaper.pdf (accessed on 29 March 2021).

25.   Zhang, Y.; Lyu, H.; Liu, Y.; Zhang, X.; Wang, Y.; Luo, J. Monitoring Depression Trend on Twitter during the COVID-19 Pandemic. *arXiv* **2020**, arXiv:2007.00228.

26.   Lu, Y.; Zheng, Q. Twitter public sentiment dynamics on cruise tourism during the COVID-19 pandemic. *Curr. Issues Tour.* **2020**, *24*, 1–7. [CrossRef]

27.   Boon-Itt, S.; Skunkan, Y. Public perception of the COVID-19 pandemic on Twitter: Sentiment analysis and topic modeling study. *JMIR Public Health Surveill.* **2020**, *6*, e21978. [CrossRef]

28.   Barkur, G.; Vibha, G.B.K. Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: Evidence from India. *Asian J. Psychiatry* **2020**, *51*, 102089. [CrossRef] [PubMed]

29.   Aljameel, S.S.; Alabbad, D.A.; Alzahrani, N.A.; Alqarni, S.M.; Alamoudi, F.A.; Babili, L.M.; Aljaafary, S.K.; Alshamrani, F.M. A Sentiment Analysis Approach to Predict an Individual's Awareness of the Precautionary Procedures to Prevent COVID-19 Outbreaks in Saudi Arabia. *Int. J. Environ. Res. Public Health* **2021**, *18*, 218. [CrossRef] [PubMed]

30.   Chakraborty, K.; Bhatia, S.; Bhattacharyya, S.; Platos, J.; Bag, R.; Hassanien, A.E. Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers—A study to show how popularity is affecting accuracy in social media. *Appl. Soft Comput.* **2020**, *97*, 106754. [CrossRef]

31.   Pastor, C.K. Sentiment Analysis of Filipinos and Effects of Extreme Community Quarantine due to Coronavirus (Covid-19) Pandemic. 2020. Available online: https://ssrn.com/abstract=3574385 (accessed on 29 March 2021).

32.   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]

33.   Kaur, S.; Kaul, P.; Zadeh, P.M. Monitoring the Dynamics of Emotions during COVID-19 Using Twitter Data. *Procedia Comput. Sci.* **2020**, *177*, 423–430. [CrossRef]

34.   Kruspe, A.; Häberle, M.; Kuhn, I.; Zhu, X.X. Cross-language sentiment analysis of European Twitter messages duringthe COVID-19 pandemic. *arXiv* **2020**, arXiv:2008.12172.

35. Dubey, A.D. Twitter Sentiment Analysis during COVID19 Outbreak. 2020. Available online: https://ssrn.com/abstract=3572023 (accessed on 29 March 2021).
36. Xue, J.; Chen, J.; Chen, C.; Zheng, C.; Li, S.; Zhu, T. Public discourse and sentiment during the COVID 19 pandemic: Using Latent Dirichlet Allocation for topic modeling on Twitter. *PLoS ONE* **2020**, *15*, e0239441. [CrossRef] [PubMed]
37. Bojanowski, P.; Grave, E.; Joulin, A.; Mikolov, T. Enriching Word Vectors with Subword Information. *CoRR* **2016**. Available online: http://xxx.lanl.gov/abs/1607.04606 (accessed on 26 March 2021).
38. Go, A.; Bhayani, R.; Huang, L. Twitter Sentiment Classification Using Distant Supervision. Available online: https://www-cs.stanford.edu/people/alecmgo/papers/TwitterDistantSupervision09.pdf (accessed on 26 March 2021).
39. Cai, M. Sentiment Analysis of Tweets using Deep Neural Architectures. In Proceedings of the 32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, QC, Canada, 3–8 December 2018; pp. 1–8.
40. Mohammad, S.M.; Bravo-Marquez, F. WASSA-2017 Shared Task on Emotion Intensity. In Proceedings of the Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA), Copenhagen, Denmark, 8 September 2017; pp. 34–39.