RESEARCH ARTICLE

ML-NLPEmot: Machine Learning-Natural Language Processing Event-Based Emotion Detection Proactive Framework Addressing Mental Health

LEILA ISMAIL[1,2,3,4], **(Member, IEEE), NADA SHAHIN**[2,3],
HUNED MATERWALA[2,3]**, ALAIN HENNEBELLE**[4]**, AND LEA FRERMANN**[4]

[1]Cloud Computing and Distributed Systems Laboratory, School of Computing and Information Systems, The University of Melbourne, Melbourne, VIC, Australia
[2]Intelligent Distributed Computing and Systems Laboratory, Department of Computer Science and Software Engineering, College of Information Technology, United Arab Emirates University, Al Ain, United Arab Emirates
[3]National Water and Energy Center, United Arab Emirates University, Al Ain, United Arab Emirates
[4]School of Computing and Information Systems, Faculty of Engineering and Information Technology, The University of Melbourne, Melbourne, VIC, Australia

Corresponding author: Leila Ismail (leila@uaeu.ac.ae)

ABSTRACT Global rapidly evolving events, e.g., COVID-19, are usually followed by countermeasures and policies. As a reaction, the public tends to express their emotions on social media platforms. Therefore, predicting emotional responses to events is critical to put a plan to avoid risky behaviors. This paper proposes a Machine Learning-Natural Language Processing-based framework to detect public emotions based on social media posts in response to specific events. It presents a precise measurement of population-level emotions which can aid governance in monitoring public response and guide it to put in place strategies such as targeted monitoring of mental health, to react to a rise in negative emotions in response to lockdowns, or information campaigns, for instance in response to elevated rates of fear in response to vaccination programs. We evaluate our framework by extracting 15,455 tweets. We annotate and categorize the emotions into 11 categories based on Plutchik's study of emotion and extract the features using a combination of Bag of Words and Term Frequency-Inverse Document Frequency. We filter 813 COVID-19 vaccine-related tweets and use them to demonstrate our framework's effectiveness. Numerical evaluation of emotions prediction using Random Forest and Logistic Regression shows that our framework predicts emotions with an accuracy up to 95.5%.

INDEX TERMS Artificial intelligence, COVID-19, emotions detection, machine learning, medical informatics, mental health, natural language processing, social media, supervised learning, vaccination.

I. INTRODUCTION

Emotions are the reflection of a person's mental state through internal signals, such as thoughts and feelings, as well as external ones, including facial expressions, behavior, and language. Despite an interest in the topic that spans back to Aristotle [1], understanding the function, origin, and mani-

The associate editor coordinating the review of this manuscript and approving it for publication was Anandakumar Haldorai.

festation of emotions is an active area of research [2], [3], [4]. Yet, understanding emotions is a central task faced by professionals across disciplines including medical and mental health specialists, social scientists, and psychologists. Understanding individuals' emotions, especially the negative ones, is crucial to avoid mental health issues and their consequences [5]. Significant positive and negative events, which can occur at any point in an individual's life, act as stimuli that derive emotions [6]. Monitoring emotions on

a societal level for instance in response to specific events, such as the introduction of policies or legislation, can help governance to respond effectively and avoid the emergence of elevated negative emotions which might lead to polarization, political disorders, or violence. In this paper, we present a methodology to monitor emotions expressed in social media in response to a specific event and evaluate our framework using case study demonstrator of the emotional response to the COVID-19 vaccination campaign in the United Arab Emirates (UAE).

The COVID-19 outbreak triggered many public emotions that have been affecting mental health. The virus started around the end of December 2019 in Wuhan, China [7], followed by its declaration as a pandemic on March 11, 2020, by the World Health Organization [7]. It is caused by Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-COV-2) and negatively impacts other vital organs such as the brain, heart, liver, pancreas, and kidney [8], [9], [10]; the virus can also cause stroke [11] and diabetes [12], [13], [14], [15]. COVID-19 infected over half a billion people and led to over 6 million deaths globally [16]. Consequently, countries have imposed precautionary measures to contain it, such as social distancing, travel bans, home confinement, business closures, and vaccination [16]. These preventive actions effectively reduced the number of infections and deaths during the pandemic [17], [18]. However, overly firm restrictions can have a negative impact on the personal level, such as loss of income, anxiety, and depression. Such measures can also affect economic and social life across the nation [19], [20], [21], [22].

A representative survey in the United States was conducted on 5,412 adults to detect mental health issues in the population in response to physical distancing and stay-at-home instructions. Researchers found that the majority showed anxiety/depression symptoms (31%), followed by stressor-related disorder symptoms (26%), and the minority had suicidal thoughts (11%) during the COVID-19 pandemic [23]. Strict lockdown restrictions have also impacted students' mental health in France. A study found that most students suffered from high levels of anxiety and stress, followed by major depression and suicidal thoughts [22]. Hence, detecting public emotions related to introducing precautionary measures or events is essential.

In this paper, we aim to detect emotions, i.e., reflections of the mental state of individuals, from their social media posts. Machine learning and natural language processing methods have been developed to automatically detect the positive, negative, or neutral sentiment of texts produced by individuals, typically in the context of a specific topic such as politics [24] or COVID-19 [25]. These approaches, however, only distinguish between broad sentiment classes and make no explicit connection to the mental state of the author. The Canadian National Research Council's Word-Emotion Association Lexicon (EmoLex) [26] relates English words to three sentiments (negative, positive, and neutral), and Plutchik's eight emotions [27], [28], [29] (joy, fear, anger, anticipation,

sadness, surprise, disgust, and trust) aiming, to connect verbal expression to the author's emotional state. Here, we use EmoLex to predict emotions expressed in posts on social media. Our prediction may inform public institutions in defining strategies to promote mechanisms and respond to public perception according to the expressed emotional state.

Many works investigated the sentiments on social media during COVID-19. Researchers studied the emotions toward COVID-19 [25], [30], [31], [32], [33], [34], [35] in general, and the COVID-19 vaccine specifically [36], [37]. Our work is distinct in two ways: first, we predict a fine-grained set of eight emotions grounded in the Psychology literature [27], [28], [29] rather than coarse (positive vs negative) sentiment. Secondly, while some work analyzed expressions of emotion in response to COVID-19 in general, it has been argued that consideration of a precise target of emotion allows for obtaining both more accurate predictions as well as more useful results [38], [39]. To the best of our knowledge, we are the first to propose a framework that captures fine-grained emotion in response to specific events, such as the event of vaccine introduction in the UAE. In other words, our framework tackles the emotional states of a population to support medical and mental health specialists, as well as social scientists, and psychologists to understand individuals' emotions, especially the negative ones associated with a specific event during a pandemic. This is with the aim to help governments develop a proactive prevention plan to avoid mental health issues and their consequences, such as awareness campaigns, measures exemptions, and surveys to detect mental health issues in individuals and act accordingly. In this paper, we present a general framework for event-specific emotion prediction with a focus on mental health impact and consider the vaccine event in the UAE as a case study. The main contributions of this work are as follows.

- We propose a proactive event-based framework that detects public emotions associated with events such as a pandemic and its related measurements.
- We implement our framework based Plutchik's 11 categories of emotions.
- We annotate a twitter dataset with Plutchik's emotions and make it publicly available for researchers.
- We evaluate the mostly used algorithms Random Forest and Logistics Regression for the prediction of tweets emotions in terms of time efficiency using accuracy, F-measure, AUC, precision, recall, ROC, and resources utilization.
- We demonstrate our framework using the COVID-19 vaccine event.

The rest of the paper is organized as follows. Section I-A reviews the related work. We describe our proposed Event-based Social Media NLP Emotion Detection (ML-NLPEmot) framework in Section II. Section III presents the UAE COVID-19 emotion detection case study demonstrator of the framework's implementation, lays out our experiments, and analyzes the results. Section IV discusses the principal

TABLE 1. Comparison between related work and this work.

Work	Country	Language(s)	Data source	Data period	Dataset size	Data Annotation Method	Algorithm(s) Sentiment analysis	Algorithm(s) Feature extraction	Data splitting	Evaluation metrics
[25]	Worldwide	English	Twitter	July 25, 2020	170,000 tweets	-	LSTM with attention layers	TF-IDF	-	Accuracy, Precision, Recall, and F-measure
[30]	The US and Brazil	English and Portuguese	Twitter	April - August 2020	English: 7,144,349 tweets Portugues: 7,125,530 tweets	Manual	Logistic Regression, Random Forest, and Linear SVM	SBert, mUSE, and FastText	-	Precision and F-measure
[31]	India	English and Indian	Twitter	March to May 2020	150,000 tweets	-	LSTM, BiLSTM, and BERT	Word2vec and GloVe	90% Training, 10% Testing	BCE, Hamming, Jaccard, and LRAP loss functions, and F- Macro and Micro
[32]	-	-	Twitter	tweets of May 30, 2020	7528 tweets	TextBlob	Random Forest, XGBoost, SVM, Extra Tree Classifier, and Decision Tree	Concatenated BoW and TF-IDF	80% train 20% test	Accuracy, Precision, Recall, and F-measure
[33]	-	-	Twitter	February to March 2020	90,000 tweets	TextBlob	SVM, Naïve Bayes, Decision Tree, Random Forest, CNN, and BiLSTM	TF-IDF, Word2Vec, GloVe, HyRank, IWV, and FastText	3 random categories	Accuracy
[34]	-	-	Twitter	Dataset 1: December 2019 - May 2020 Dataset 2: January - March 2020	Dataset 1: 23,000 tweets Dataset 2: 226,668 tweets	VADER	Proposed Fuzzy Rule Model	Bag-of-words and Doc2Vec	-	Sensitivity, Precision, Recall, F-measure, Log Loss, and Mean Absolute Error
[35]	-	English	Twitter	23 March 2020-10 May 2021	31,233 tweets	TextBlob	SVM, Logistic Regression, Random Forest, Multi Naïve Bayes, LSTM	-	75% train 25% test	Accuracy, Precision, Recall, and F-measure
[36]	Philippines	English and Tagalog	Twitter	1-31 March 2021	993 tweets	Manual	Naïve Bayes	TF-IDF	10-fold cross-validation	Accuracy, precision, and recall
[37]	-	English	Twitter	-	25,004 tweets	TextBlob	Random Forest, Decision Tree, Logistic Regression, SVM	-	70% train 30% test	Accuracy, F-measure, precision, and recall
[40]	-	English	Twitter	July – Sept 2021	113,969 tweets	-	Graph Convolutional Neural Network, ALBERT, ALBERT + LSTM	LIWC	10-fold cross-validation	Precision, Recall, and F-measure
[41]	-	English	Twitter	March 2008 – October 2021	3,654,544 tweets	previously collected annotator-rated emotions	Bert	-	80% train 20% test	F-measure
[44]	Indonesia	Indonesian	Twitter	-	4,401 tweets	Previously annotated	BiLSTM, SVM, Logistic Regression	Word2Vec, FastText	10-fold cross-validation	Accuracy, Precision, Recall, and F-measure
This study	**UAE**	**English**	**Twitter**	**Jan 29, 2020 – July 31, 2022**	**15,455 tweets**	**Manual**	**Random Forest and Logistic Regression**	**Concatenate BoW and TF-IDF**	**10-fold cross-validation**	**Accuracy, precision, recall, F-measure, ROC curve, AUC (Area Under the Curve), and execution time**

TABLE 2. Emotions considered in the literature and this study.

Work	Emotions											Emotions Framework
	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust	Neutral	Negative	Positive	
[25]	✓	✗	✗	✓	✓	✓	✗	✗	✗	✗	✗	✗
[30]	✓	✗	✗	✓	✓	✓	✗	✗	✗	✗	✗	✗
[31]	✗	✗	✗	✓	✗	✓	✗	✗	✗	✓	✓	✗
[32]	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	✓	✗
[33]	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	✓	✗
[34]	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	✓	✗
[35]	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	✓	✗
[36]	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	✓	✗
[37]	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	✓	✗
[40]	✓	✗	✓	✓	✓	✓	✓	✗	✓	✗	✗	✗
[41]	✓	✗	✓	✓	✓	✓	✓	✗	✗	✗	✗	✗
[44]	✓	✗	✗	✓	✓	✓	✗	✗	✗	✗	✗	✗
This study	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

findings and lessons learned. Finally, we conclude the article with future research directions in Section V.

A. RELATED WORKS

Table 1 presents the works in the literature which focused on predicting emotions and/or sentiments on social media using machine learning [25], [30], [31], [32], [33], [34], [35], [36], [37], [40], [41]. This is based on the properties of the considered datasets, the emotion detection algorithms used, and the annotation strategy. As shown in the table, to annotate the dataset, [32], [33], [35], [37] used TextBlob [42], and [34] used VADER [43], without manual validation. Furthermore [30], [36] annotated the data manually. However, [25], [31], and [40] do not report on their annotation method.

Table 2 presents the emotions and sentiments considered each work. Most of these works [31], [32], [33], [34], [35], [36], [37] focused on detecting the sentiments (negative, neutral, and/or positive). The rest of the works focused on a few emotions, mainly Fear, Sadness and Anger [25], [30], [31], [40], [41], [44], [45].

Table 2 shows that no work which considers fine-grained set of 8 emotions and 3 sentiments grounded in the psychology literature, such as Plutchik's [2], [27], [28], [29]. In this paper, we address this void. In addition, we propose an event-based framework for emotions' detection, along with manual annotation.

II. METHODOLOGY

A. PROPOSED ML-NLPEMOT FRAMEWORK

To detect event-specific emotional states of a population using social media posts, we present a machine learning model-building methodology and evaluation strategy. We discuss challenges in emotion detection and propose a solution that we test in the context of pandemic containment events. Public health professionals, policymakers, and mental health epidemiology researchers can use the framework as a step-by-step guide to detect population emotions on social media whenever an event is introduced by policy decision-makers. They can use our data repository [46] and machine learning models that we developed for emotion detection. Fig. 1 shows our proposed framework. It consists of eight stages: 1) domain understanding, 2) data collection, 3) data pre-processing, 4) data annotation, 5) feature extraction, 6) model building, 7) model evaluation, and 8) model deployment.

Stage 1: Domain Understanding

The first step in analyzing emotions is to define the problem by specifying the event(s) to be studied in a particular country. Additionally, the language(s) must be specified because the population can express themselves in different languages. Finally, the platform from which the observations (i.e., social media posts) will be extracted must be specified to limit the search. Consequently, the corresponding observations should be extracted based on the time of the event(s) during which it occurred as stated in Equation (1).

$$P_{L,c} = \sum e_{start} - e_{end} \qquad (1)$$

where $P_{L,c}$ represents the social media posts in a specific language (L) for a specific country (c), e_{start} and e_{end} represent the event's start and end respectively.

Stage 2: Data collection

There are multiple ways to data collection. Typically, data are obtained by social media platforms' data access APIs, but it may be possible to request it directly from a social media platform or a database owner. During this step, the period related to the event(s) under study must be specified. In addition, a set of keywords related to the event(s) under study must be determined, to extract all and only relevant social media posts. Keywords can be designed in collaboration with domain experts or taken from the prior literature.
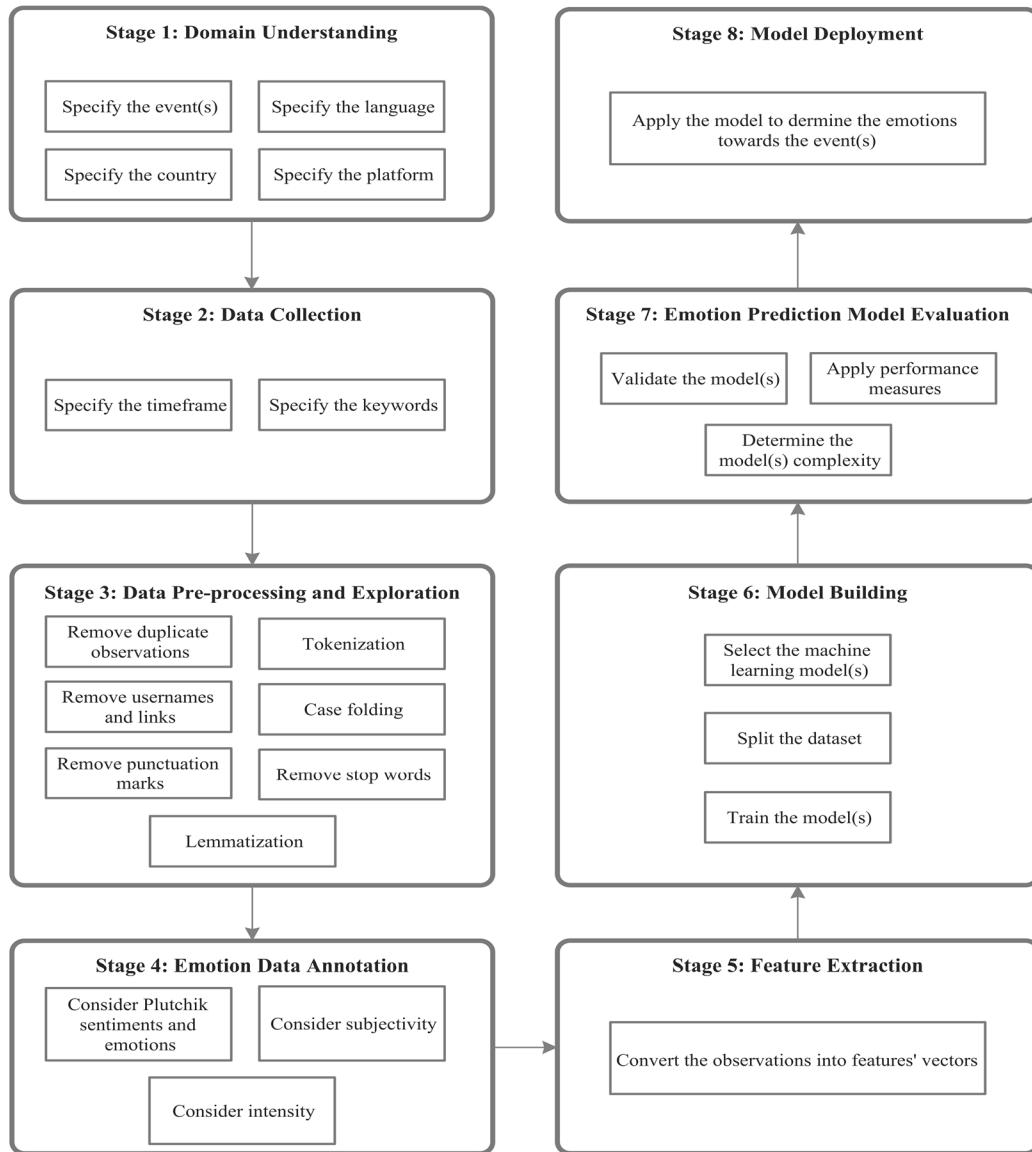
Social media data can be extracted to a local database or a cloud [47], [48], [49]

Stage 3: Data pre-processing and exploration

Following common practice in natural language processing, we apply the following pre-processing steps to normalize the text dataset and remove sensitive or irrelevant information, after which the dataset is explored to extract useful insights.

Removing duplicated observations helps produce accurate results regarding the emotions and model performance. Keeping the duplicates would imply a change from the natural data distribution and hence worsen generalization to unseen test cases.

Removing usernames and links when an observation includes a username that indicates the author's identity.

Usernames denoted by '@username,' should be removed to protect user privacy, and as the name translates to no emotions. Similarly, any hyperlinks in the observations are deleted.

Removing punctuation marks such as [", #, $,:, ?, @] from the observation, as they do not carry lexical semantic information about emotions, and can hinder the performance of machine learning algorithms.

Tokenization divides the raw text into atomic chunks, or tokens [50]. Paragraphs are tokenized into sentences, sentences are tokenized into words, and punctuation is stripped from words. Tokenization helps to interpret the meaning of observations by analyzing the sequence of words.

Case folding is the process of converting all the characters in the text to either upper case or lower case. Therefore,
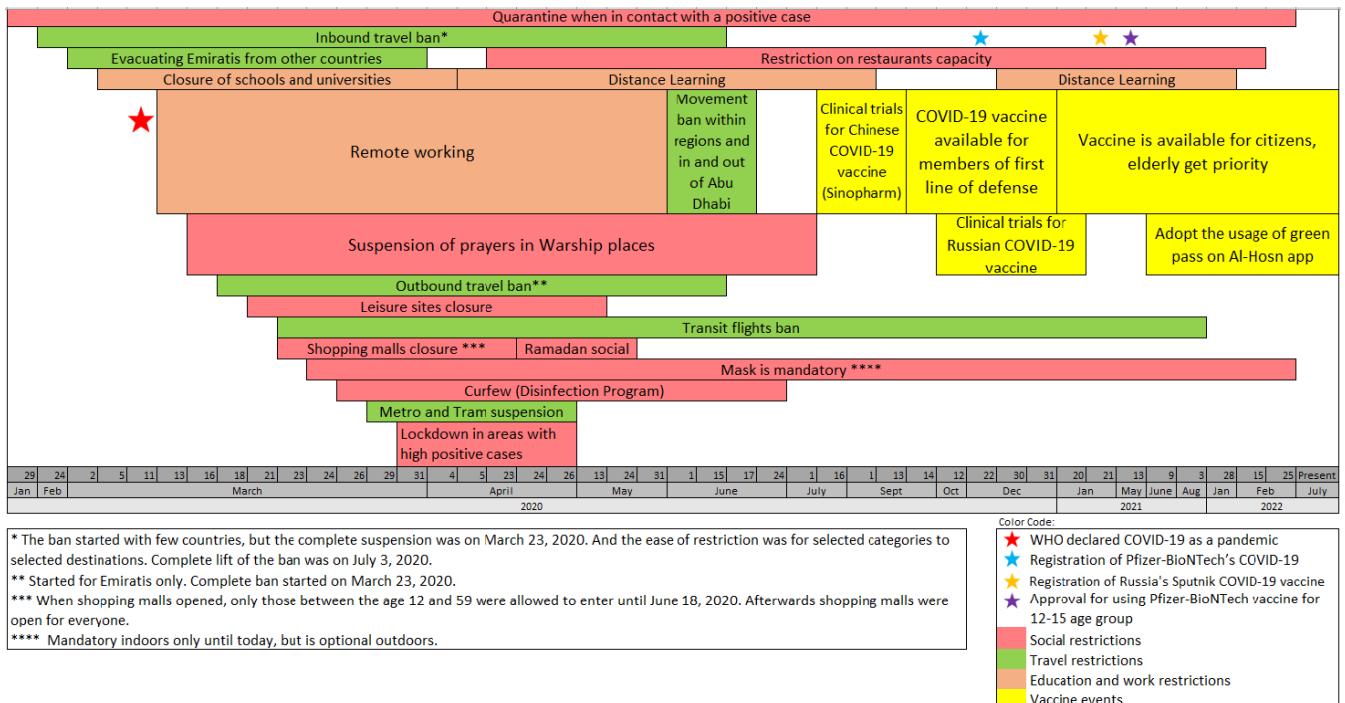
FIGURE 2. Main precautionary measures and restrictions to tackle COVID-19 in the UAE.

instances of the same word written in different capitalization formats, such as 'Corona,' 'CORONA,' and 'corona,' are now treated as identical by the machine learning approaches.

Removing stop words eliminates purely functional words so that the algorithm focuses on the content words that convey valuable information. Stop words are frequent and do not carry meaning in themselves. Examples of stop words in the English language are ['a', 'an', 'at', 'of', 'the'].

Lemmatization translates an observed word form into its root form or lemma based on its meaning and context [51]. For instance, 'viruses' and 'virus' would be lemmatized to the root form 'virus' by eliminating inflectional endings. It enables the machine learning algorithm to consider different inflected forms of a word as a single word.

Stage 4: Emotion data annotation

To (i) evaluate the quality of predicted emotions against a gold standard dataset, and (ii) train supervised emotion classifiers on ground truth data, we label observations with the emotion(s) they express. While unsupervised approaches, which do not rely on labeled training data could be an alternative, without guidance from human labels, the algorithm might misinterpret the tasks or ambiguous words in the input. In the emotions data annotation step, the retrieved observations are annotated based on Plutchik sentiments (positive, negative, and neutral) and emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, and trust) while considering each word's subjectivity and intensity. Subjectivity refers to the opinions and judgment of the information by the author, bringing insights into the associated emotion. Intensity quantifies the impact of modifiers on the whole sentence, and thus

the associated emotion. For example, "good" implies less intensity than the expression "very good".

Stage 5: Feature extraction

The pre-processed observations (i.e., posts represented as lists of words) are converted into numerical feature vectors which can be processed by the machine learning algorithms. To convert each word in observation to a numerical value, feature extraction techniques such as TF-IDF, Pointwise Mutual Information, Word2Vec, BoW, and Ranking [52]. Each observation is then represented as a vector of features, and the similarity of two observations can be quantified using standard vector similarity measures such as Cosine similarity.

Stage 6: Model building

In this stage, the machine learning model must be selected based on established methodologies in the related works, expert advice, or by exhaustive evaluations of several models and selecting the best-performing one. A labeled dataset must be split into a training and testing sets, for example using a fixed split (e.g., 70% training data and 30% testing data). K-fold cross-validation [53] is an attractive alternative, where the dataset is split K times into different training and testing sets and trained/tested with each split once. This is particularly useful in the face of small datasets, as every observation will be used for training and testing, while ensuring no overlap between the two sets. The classification model will be trained using the training dataset, and its performance will be evaluated on the test set.

Stage 7: Emotion prediction model evaluation

Model evaluation concerns validating the classification model using the testing dataset. In this step, the model is

evaluated based on several performance metrics such as accuracy, precision, recall, F-measure, and AUC. Accuracy is the most widely used evaluation metric [54]. Lastly, the execution time should be calculated to determine the model's complexity.

Stage 8: Model deployment

Following satisfactory evaluation results in Stage 7, in this final stage of the framework, the developed model can be applied to determine residents' emotions toward the specified event(s) in Stage 1. The model should be retrained to consider future social media posts.

B. THE UNITED ARAB EMIRATES (UAE) COVID-19 EMOTIONS DETECTION CASE STUDY

Vaccination was introduced as a step toward containing the COVID-19 outbreak. However, the proliferation of vaccines, the absence of data about the short- and long-term side effects, and their efficacy against the virus and its evolving mutation drove the explosion of different emotions, such as fear, anticipation, and anger, that could lead to mental health issues. Our primary goal, in this case study, is to understand the emotional response of UAE residents to COVID-19 vaccines and to evaluate the performance of our machine learning event-based emotion detection framework by considering Tweets about the vaccines to gauge the impact on individuals' emotions. Fig 2 summarizes the data that we extracted from WHO [7] about the declaration of the pandemic, and the news posts about precautionary measures adopted by the UAE, as announced by the National Emergency Crisis and Disasters Management Authority [16], to tackle the pandemic since January 29, 2020. In what follows, we will refer to precautionary measures as events. The events related to vaccines started on July 16, 2020. In this case study, we explore emotions related to COVID-19 in the UAE in general. In addition, we evaluate our framework to detect emotions related to vaccine events through different performance measures.

1) COVID-19 VACCINE-RELATED PRECAUTIONARY MEASURES

To cope with the increasing COVID-19 infections in UAE [55], [56], [57], on July 16, 2020, the government announced a call for volunteers to undergo clinical trials for potential COVID-19 vaccines. Then, on September 14, it announced the availability of the Sinopharm COVID-19 vaccine for the first line of defense members. The following month, the government announced clinical trials of a Russian adenovirus-based vaccine (Sputnik). Moreover, in December 2020, there was an official registration of Pfizer-BioNTech's COVID-19 vaccine, with priority given to the elderly.

Towards the end of January 2021, the government announced the eligibility of people aged 16 and above to take the vaccine and the requirements of a Polymerase Chain Reaction (PCR) test every seven days for federal government employees who did not receive the two doses of the vaccine. The Ministry of Health dedicated vaccination centers to the

TABLE 3. Keywords used to extract data from twitter.

	Keywords
COVID-19	Coronavirus, Corona, Covid, COVID-19, Sarscov2, ncov2019, Lockdown, Social distance, Quarantine, Stay home, Work from home, Pandemic, Outbreak, Travel ban, Travel restriction
Vaccination	Vaccination, Covid vaccine, Vacc, Vax, Sinopharm, Pfizer-BioNTech, Pfizer, Sputnik V, Sputnik, AstraZeneca
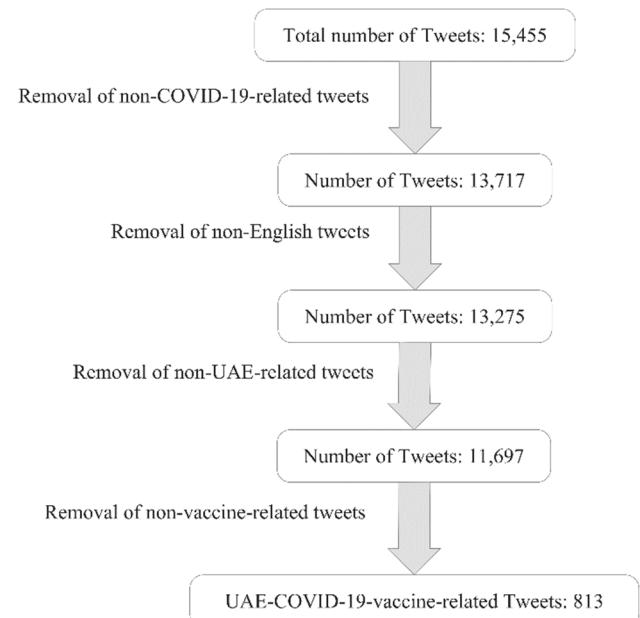
FIGURE 3. Tweets filtering steps.

elderly and people with chronic diseases in February. As of July 2022, more than 24 million doses have been administered.

III. RESULTS

This section explains how we used our ML- SMEmot framework to analyze the emotional response to COVID-19 vaccination in the UAE as expressed in social media.

Stage 1: Domain Understanding

In this case study, we are interested in collecting English tweets regarding the COVID-19 vaccine-related events in the UAE. Given a population consisting of over 88% expatriates of more than 200 nationalities [58], English is the most common language in UAE [59].

Stage 2: Data collection

We collect tweets posted between 29-01-2020 and 31-07-2022. We divide the duration into three periods: 1) 29-01-2020 – 14-09-2020, from the first case of COVID-19 in UAE until the vaccination announcement, 2) 14-09-2020 – 09-12-2020, from the announcement of vaccination until the commencement of the vaccination program, and 3) 09-12-2020 – 31-07-2022, from the commencement of

the vaccination program until the time of writing this paper. Additionally, we use the keywords in Table 3 to extract the observations using the Snscrape package in Python 3.9. The keywords consist of the ones related to COVID-19 in general and the ones related to vaccination.

Stage 3: Data pre-processing and exploration

We remove duplicate tweets which is posted by the same user ID. However, we kept duplicate tweets posted by different users, i.e., retweets. This is because retweets would emphasize on embedded emotions. For each tweet, we retain the date created and the text, while we remove other irrelevant attributes such as username, retweets count, number of likes and comments. Additionally, we case-folded the tweets to lowercase and removed the word 'RT' from the tweets, which indicates retweet. Fig. 3 illustrates the steps we took to filter the UAE-COVID-19-vaccine-related tweets. First, we extracted a total of 15,455 tweets that contained at least one of our COVID-19-related keywords, then we manually removed the non-English tweets, which are the ones that include a mix of multiple languages. Furthermore, we manually removed the non-UAE-related tweets which were about other countries' COVID-19 precautionary measures or situations that were not related in any way to the UAE's procedure for containing the virus. Finally, we filtered the vaccine-related tweets based on the 25 keywords mentioned in Table 3, which resulted in a final dataset of 813 UAE-related tweets about the event of vaccination.

Table 5 in Appendix A shows the output for the above-mentioned pre-processing steps on a sample observation taken from our dataset.

A. EXPLORATORY DATA ANALYSIS

Fig. 4 illustrates the number of positive COVID-19 cases on the left y-axis, versus the number of vaccine-related tweets, on the right y-axis each month. The figure shows that in most cases, the changes in positive cases are in line with the changes in vaccine-related tweets. However, in July 2022, we noticed a sudden increase in vaccine-related tweets despite the low positive cases due to the ease of restrictions in the UAE. We hypothesize that the population was getting more aware of the need for the vaccine and booster shots. Fig. 5 shows counts of unigrams (single words), bigrams (word pairs), and trigrams (sequences of 3 words). These figures show a dominance of keywords (or key phrases) related to COVID-19 and the vaccine. Fig. 5(a) shows that 'Pfizer' is tweeted about most, followed by Sinopharm. The reason behind this could be that the Pfizer vaccine is produced by a well-known company or because the success rate of Pfizer exceeded Sinopharm. Fig. 5(c) shows that the tweets related to Pfizer were mostly about the shots taken. Both Dubai (Fig. 5(a)) and Abu Dhabi (Fig. 5(b)) feature frequently in the Tweets. Fig. 5(c) suggests that this is because of the travel restrictions between both cities which impacted daily life on the personal level and nationwide economical level.

Fig. 6 and 7 illustrate COVID-19-word clouds. To give insights into the tweets' context during the pandemic,

we removed the COVID-19-related terms such as "Covid," "Covid-19," "Coronavirus," "Vaccine," and "Vaccines". The reason behind such a decision is that all word clouds represent COVID-19-related tweets; hence, such terms will be the most frequent. However, we did not remove the terms related to the different types of vaccines to find the correlation between each emotion and the vaccine types. Fig. 6(a) represents a word cloud for non-vaccine-(but COVID)-related tweets. People in the UAE mostly tweeted about quarantine and lockdown. They also tweeted about their daily routine using the word "day". Fig. 6(b) represents a word cloud for vaccine-related tweets. China was the most frequent word since the primary vaccine used in the UAE is the Chinese vaccine, Sinopharm. Fig 7 represents the word clouds for the tweets before and after the introduction of the Sinopharm vaccine on 14 September 2020 [59]. Before the vaccine was introduced, people frequently tweeted about other COVID-related news using the hashtag "NewsBits". On the contrary, when vaccination started, Pfizer dominated the social media discourse.

Stage 4: Emotion data annotation

For every tweet, we used Python 3.9 TextBlob 0.17.1 [42], vaderSentiment 3.3.2 [43], and NRCLex 4.0 [60] libraries. The latter is based on EmoLex [26]. TextBlob and VADER produce the tweets' sentiment (negative, positive, or neutral), and EmoLext provided Plutchik's eight emotions (joy, fear, anger, anticipation, sadness, surprise, disgust, trust) and two general sentiments (negative and positive). We automatically tag tweets with a single label using EmoLex emotion if and only if all three libraries agreed on one of the sentiment classes (negative or positive). In all other cases, the authors L.I., N.S., H.M., and A.H. manually single labeled the tweet based on EmoLex with full agreement. In our manual annotations, we also consider the level of the subjectivity of the tweet and its emotional intensity in terms of the existing modifiers. Fig 8 shows the low rate of agreement of all three methods, implying that most tweets were manually annotated with fine-grained emotions. We added a "neutral" class in our manual annotation phase to indicate tweets that are void of emotion.

In summary, we annotated each tweet with the emotions it expressed in terms of three sentiments (positive, negative, and neutral) and Plutchik's eight emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, and trust) [2]. We considered the tweet's author's opinion and judgment (subjectivity), and its intensity. The latter depicts the impact of the modifiers on each tweet since it would affect the classification of a tweet into one of the eleven classes (i.e., Plutchik's sentiments and emotions) [26].

B. EMOTION-BASED ANNOTATED DATA EXPLORATORY ANALYSIS

Fig. 9 represents the prevalence of different emotions in our manually annotated tweets on UAE-COVID-19 for both vaccine-related (red) and non-related (blue) tweets. Most of
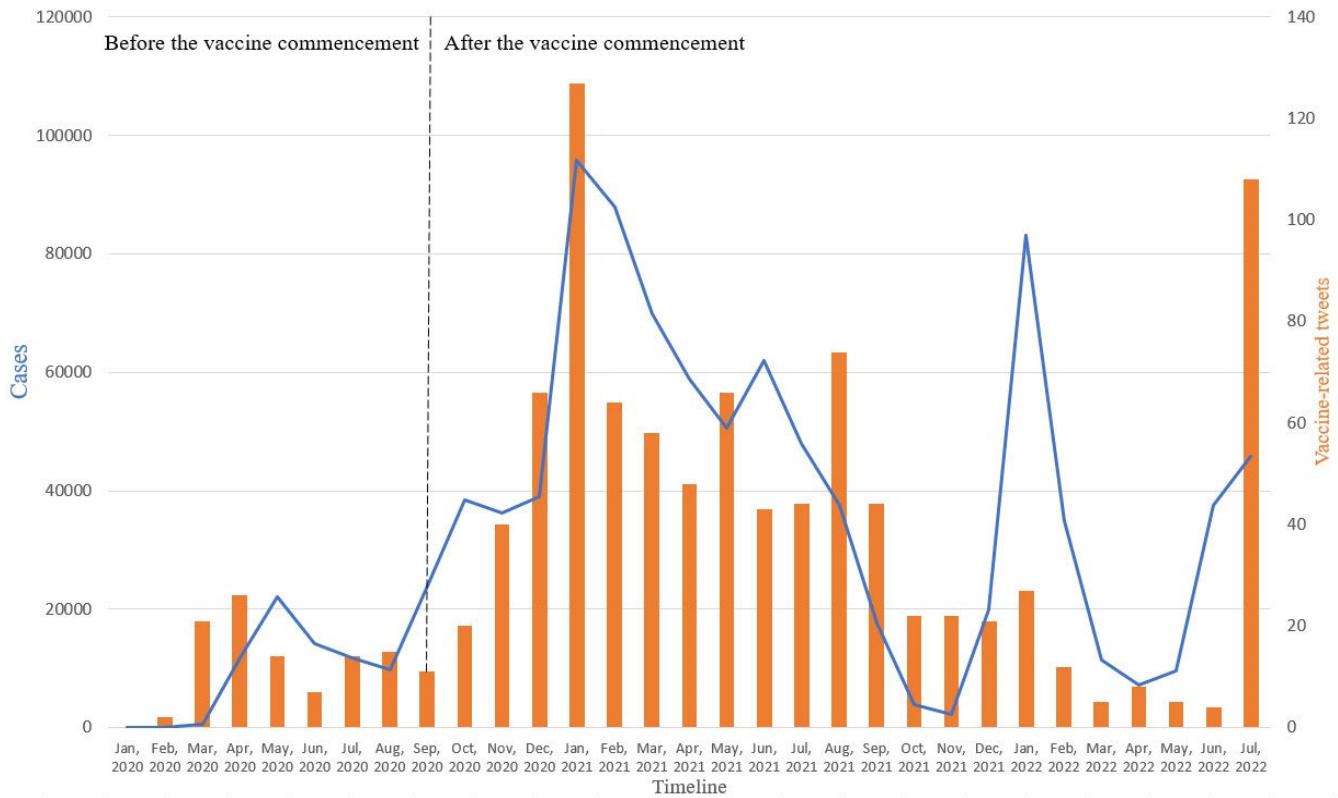
FIGURE 4. The number of positive cases versus the number of vaccine-related tweets each month.

(a) Non-vaccine related tweets   (b) Vaccine-related tweets
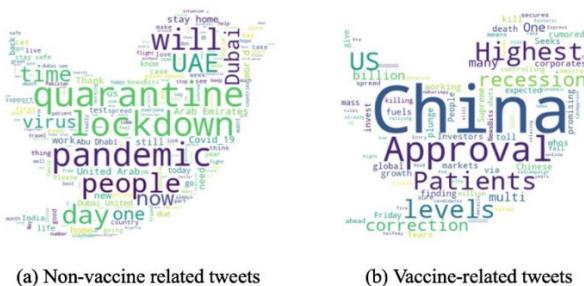
FIGURE 5. Word clouds for tweets before and after the Sinopharm vaccine.

the tweets are classified as positive in both categories. Negative emotions are the second most prevalent in the non-related category, while trust is the second most prevalent emotion (after positive) in the vaccine-related category. Neutral is the third most common emotion in COVID-related tweets, while fear is the third most prevalent in vaccine-related tweets. Fig. 10 presents the emotions based on manual annotation before and after the vaccine commencement (September 14, 2020). People in the UAE tweeted more positive and fewer negative tweets after the vaccine when compared to the tweets before the vaccine. We also find that anger increased after the introduction of the vaccine, while trust declined slightly. Overall, the manual annotation suggests that the UAE's pop-

ulation's emotions were dominated by positive responses throughout. Before the vaccine, people in the UAE stayed mainly positive, with high trust in the government's decisions. However, although positive emotions dominated after the vaccine, fear of the vaccine's side effects was expressed. A relaxation of restrictions may account for additional fear among parts of the population. Overall, positive and trust emotions dominate in the tweets, suggesting broad support of government decisions among social media users.

Fig. 11 represents a word cloud for each emotion based on the manual annotation. We can see that people in the UAE mention Pfizer in tweets with positive sentiments, surprise, and anticipation emotions. The UAE, Dubai, and the capital Abu-Dhabi mainly were mentioned in tweets with positive, neutral, and trust emotions. People tweets express trust in the government's decisions in handling the pandemic and the vaccination process but also tweet about the cities of Abu Dhabi and Dubai with anger. This is due to lockdowns, and movement restrictions between the two cities. Both sad and fearful tweets focus on travel bans and flight suspensions between UAE and India, as Indians could not travel to their home country during the interruption period.

Stage 5: Feature extraction

After annotation of tweets with gold emotion labels, we transform each tweet's text into a feature vector, represented by numeric features, which is accepted as input by the machine learning algorithms that we will apply to predict

(a) Unigram term count

(b) Bigram term count

(c) Trigram term count
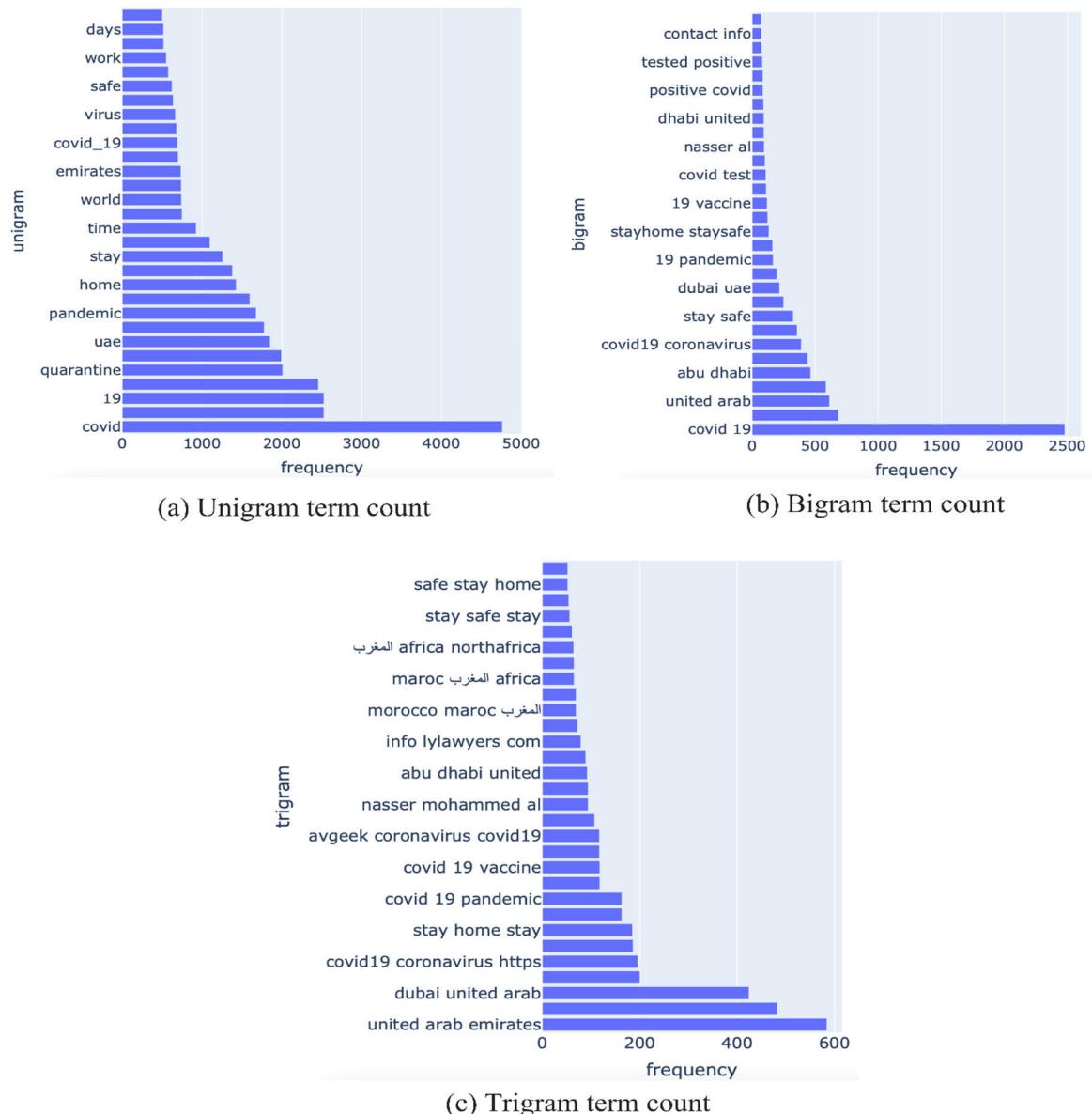
FIGURE 6. Term count (unigram, bigram, and trigram) for vaccine-related tweets.

(a) Tweets before the vaccine

(b) Tweets after the vaccine
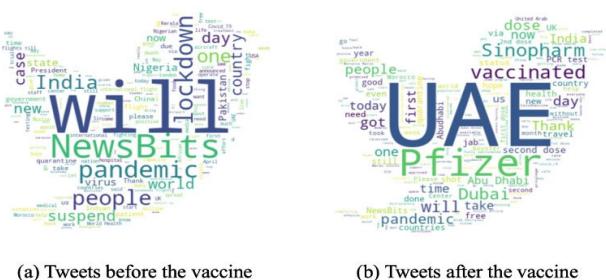
FIGURE 7. Word cloud for non-vaccine-related and vaccine-related tweets.

tweet emotion classes [61]. We use the concatenated Bag of Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF) representations as numeric features. This method is the most accurate for sentiment analysis in the literature and performs better than BoW and TF-IDF alone [32]. The BoW is a technique that calculates the frequency of each word in a tweet regardless of its order, where the frequency of a word represents its importance [52] and is used as a feature [62]. It consists of two steps: 1) establishing the dimensions of the feature vector as the vocabulary of unique words in all the observations and 2) creating a vector for each observation in the dataset by instantiating each vector element with the term frequency of a particular word in the corresponding observation. An example of BoW is available in Appendix B.

TF-IDF is a statistical technique that depicts how vital the word is to a tweet in a collection of tweets [63]. It removes the limitation of BoW, where the most common word is given
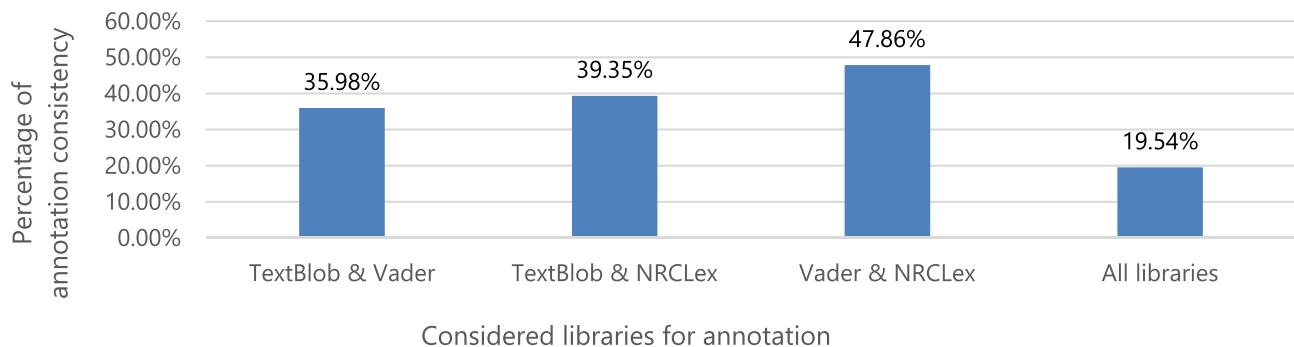
FIGURE 8. Percentage of Annotation Consistency between TextBlob, Vader, and NRCLex libraries.
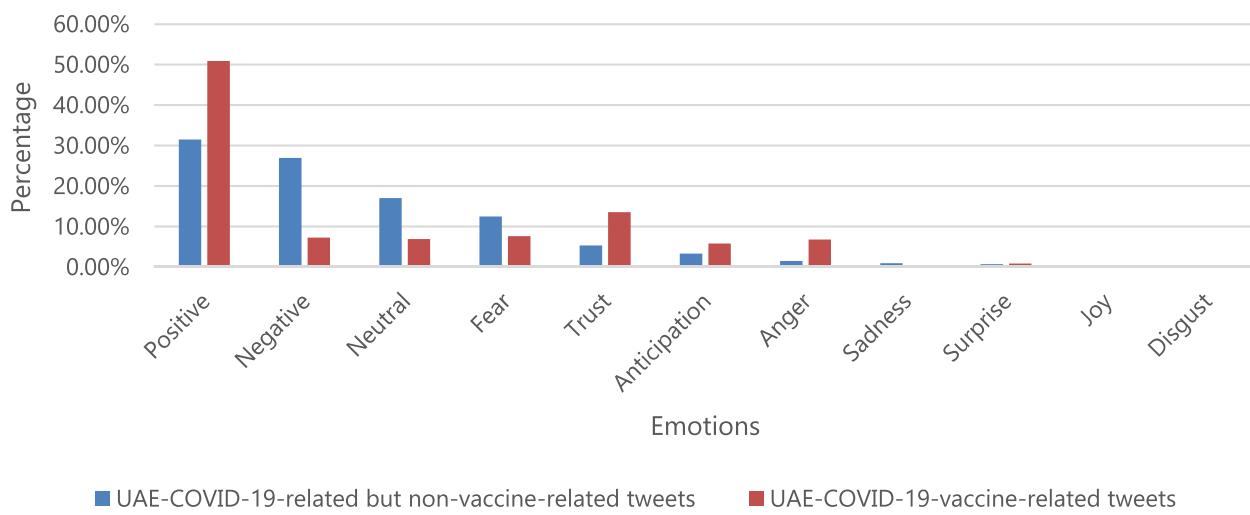
FIGURE 9. Manually annotated emotions for UAE-COVID-19-related vaccine-related and non-vaccine-related tweets. The emotions Joy and Disgust were presented in the non-vaccine related but not in the vaccine-related tweets.
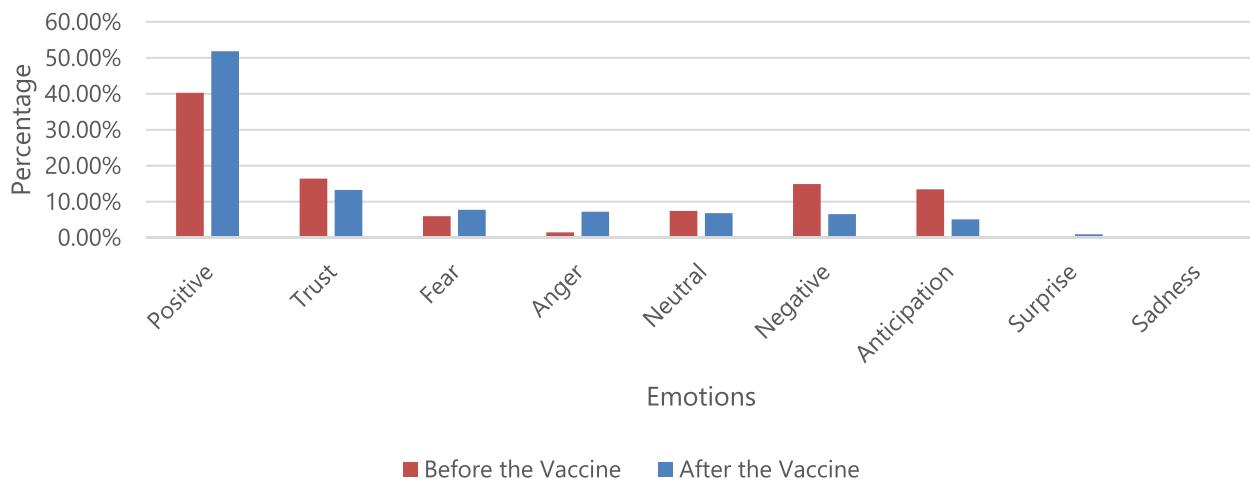

FIGURE 10. Tweets emotions before and after the vaccine based on manual annotation.

more importance. Instead, this technique compares the word's frequency in a targeted tweet against the word's average frequency in the whole collection. Intuitively, a word that is frequent in the target tweet but infrequent in the collection overall is an important feature as it is distinctive for the tweet at hand. Words that are frequent overall (e.g., "the", or "COVID") are less distinctive and receive a lower TF-IDF score. E.g., unlike the word "COVID" which is very

(a) Positive Tweets

(b) Neutral Tweets

(c) Negative Tweets

(d) Trust Tweets

(e) Surprise Tweets

(f) Anger Tweets

(g) Anticipation Tweets

(h) Sadness Tweets

(i) Fear Tweets

FIGURE 11. Word clouds for UAE-COVID-19 tweets based on manual annotation of emotion.
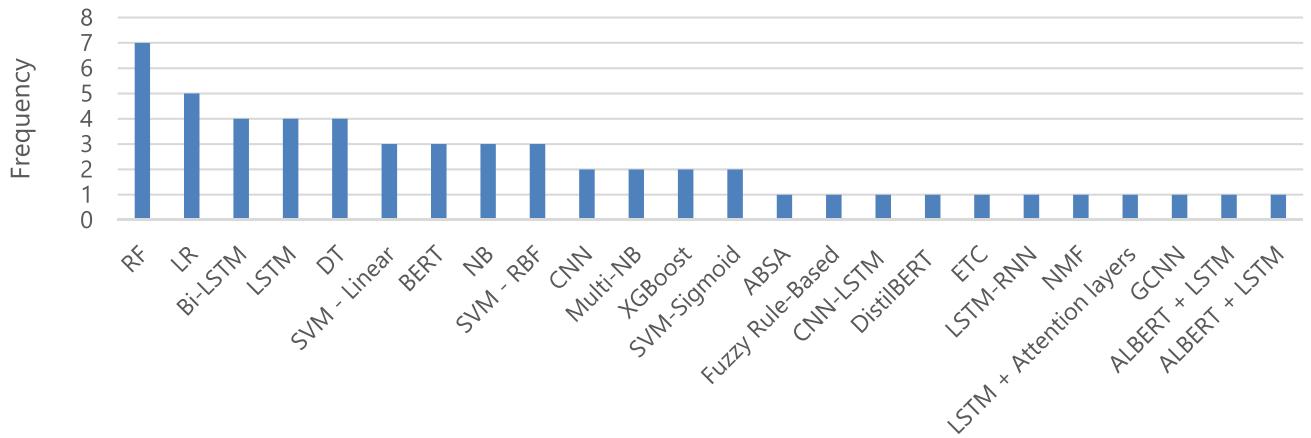
FIGURE 12. Frequency of classification models used in literature for COVID-19 sentiment analysis.
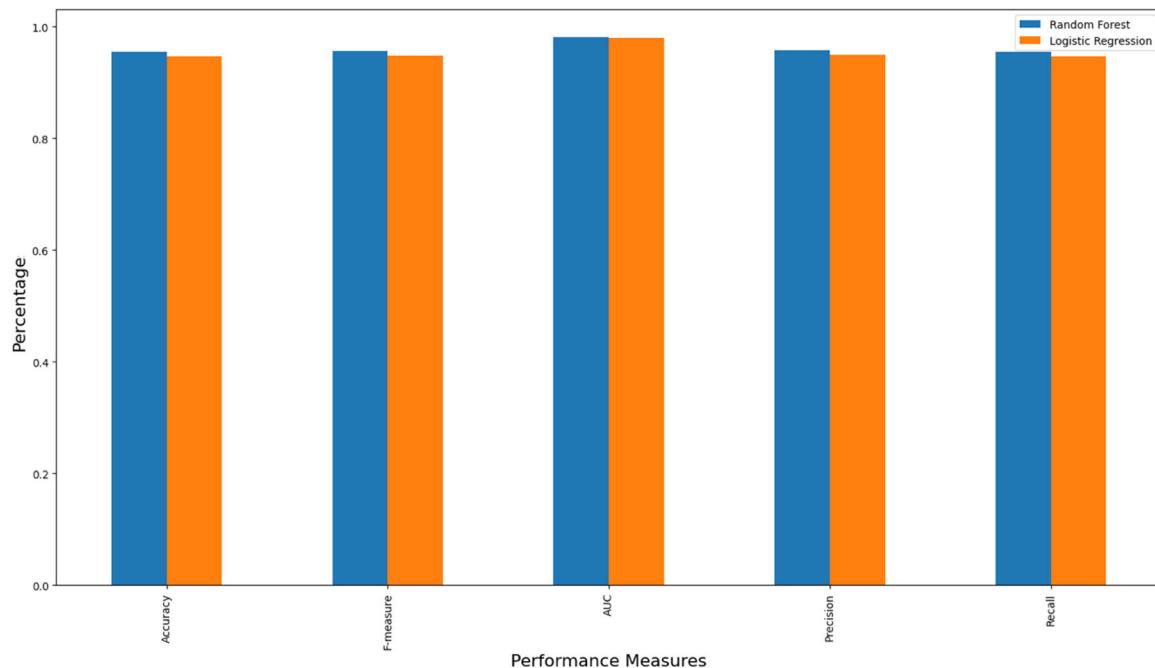
frequent, the word "Pfizer" might be mentioned in only a small subset of tweets and is hence considered distinctive. If we aim to represent tweets related to the vaccine, the word "Pfizer" would receive a higher TF-IDF weight compared to the word "COVID" [52]. In a tweet, a word that has higher concatenated BoW and TF-IDF score represents a stronger

TABLE 4. Hyperparameter tuning for the machine learning algorithms under study.

Algorithm	Hyperparameters	Optimal value(s) used in the literature	Value(s) used in our experiments
	Number of estimators/trees	300 [32], 600 [25], NR [30], [33], [37], [45]	Range from 300 to 600 with 10 intervals. (310*)
	Splitting criteria	NR [25], [30], [32], [33], [37], [45]	Gini* & Entropy
	Maximum tree depth	300 [32], [33], 50 [25], NR [30], [37], [45]	Range from 50 to 300 with 10 intervals. (90*)
	Minimum number of samples	2,5*,10 [25], NR [30], [32], [33], [45]	5
	Minimum number of samples at a leaf node	NR [25], [30], [32], [33], [37], [45]	1
	Minimum weighted fraction of the total of weights at a leaf node	NR [25], [30], [32], [33], [37], [45]	0
RF	Maximum number of features	NR [25], [30], [32], [33], [37], [45]	None*, sqrt, and log2
	Maximum leaf node	NR [25], [30], [32], [33], [37], [45]	None (Unlimited)
	Minimum impurity	NR [25], [30], [32], [33], [37], [45]	0
	Bootstrap samples usage	NR [25], [30], [32], [33], [37], [45]	True
	Out-of-bag samples usage	NR [25], [30], [32], [33], [37], [45]	False
	Number of jobs to run in parallel	NR [25], [30], [32], [33], [37], [45]	-1 (all available cores are used)
	Random state	27 [32], NR [25], [30], [33], [37], [45]	27
	Verbose	NR [25], [30], [32], [33], [37], [45]	0
	Class weight	NR [25], [30], [32], [33], [37], [45]	Balanced_subsample
LR	Solver	lbfgs [35], NR [25], [30], [37], [44]	lbfgs
	Max iteration	1000 [35], NR [25], [30], [37], [44]	Range from 1 to 1000 with 100 interval (100*)

NR: Not Reported; *: Optimal value

FIGURE 13. Performance measures for the emotion prediction models.

relationship with the corresponding Plutchik's emotion or sentiment than the words having lower scores. This relationship is learned by the machine learning algorithm during the model building stage.

The TF-IDF for each unique word in each tweet is calculated using Equation (2).

$$TF - IDF_{w,i} = TF_{w,i} \times IDF_w \qquad (2)$$
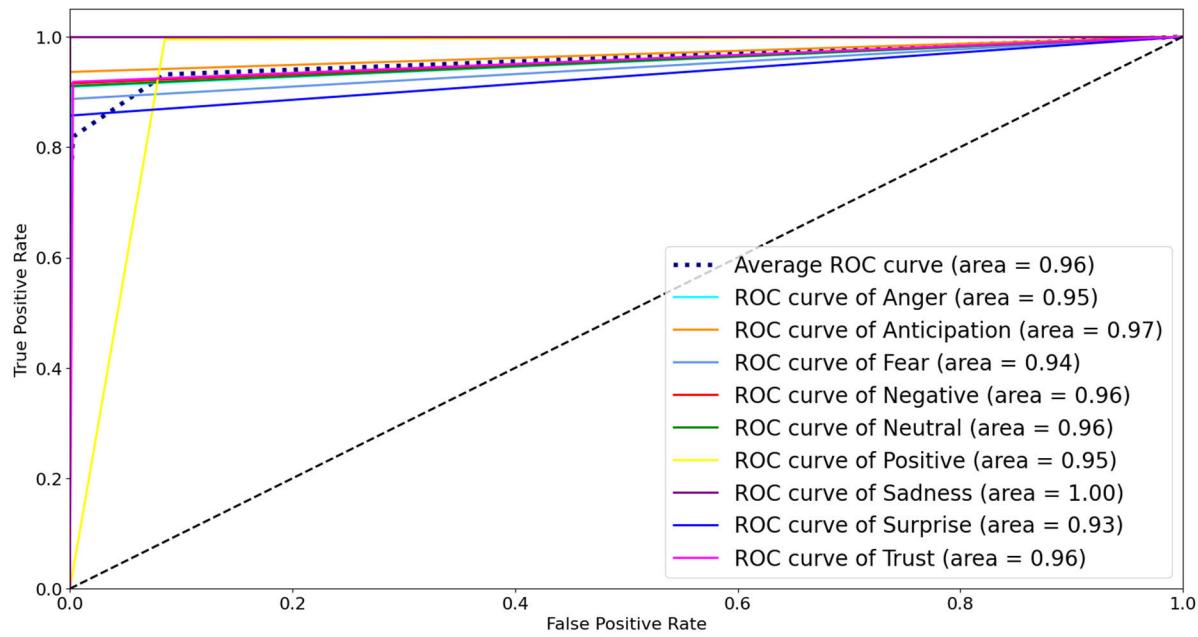
FIGURE 14. ROC curves for Random Forest model.
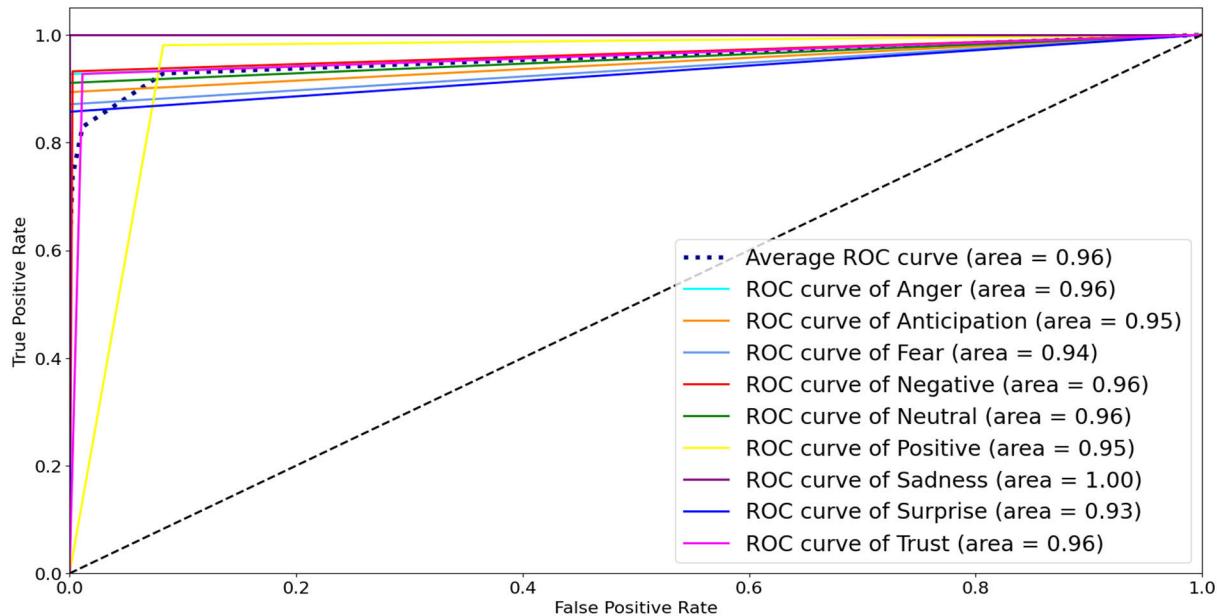
FIGURE 15. ROC curves for Logistic Regression model.

where $TF-IDF_{w,i}$ is the TF-IDF for the word w in the ith tweet. $TF_{w,i}$ is the term frequency of the word w in the ith tweet, and is calculated by dividing the number of times a word is repeated in a tweet by the total number of words in the same tweet (Equation 3).

IDF_w is the inverse document frequency for the word, i.e., the reciprocal of the number of documents that contain w. IDF_w is calculated by squashing the division of the total number of tweets in the dataset by the number of tweets

that includes the specific word through the $\log_{10}$ function (Equation 4). The log function is a convenient way to handle a large number of tweets and dampen an overly high impact of extremely high values [52]. In this formula, terms common in each tweet in the dataset, such as "the" or "and," will rank low according to TF-IDF. However, words that frequently appear in a tweet, while not appearing many in other tweets, would mean that such words are relevant, i.e., the word "China" will rank high and is considered relevant
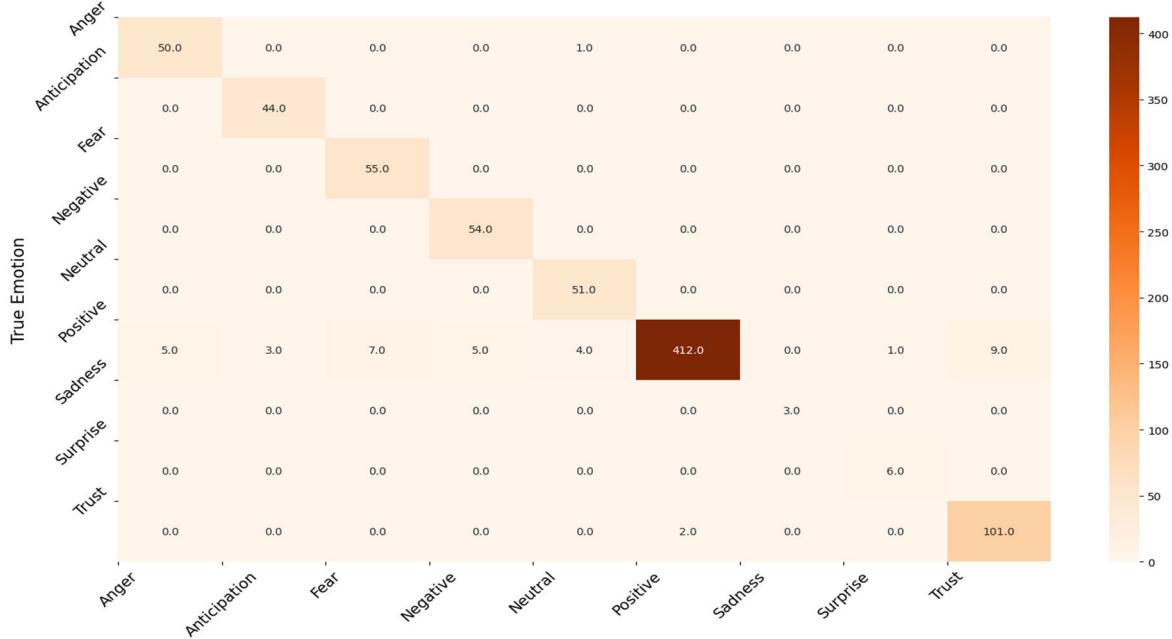
**FIGURE 16.** Confusion matrix for Random Forest model.

FIGURE 17. Confusion matrix for Logistic Regression model.

to the topic "vaccine". If we aim to find tweets related to the Chinese vaccine, the word "China" would be highly weighted compared to the word "the" [52].

$$TF_{w,i} = \frac{n_{w,i}}{NE_i} \qquad (3)$$

$$IDF_w = \log_{10} \frac{N}{\sum_{i=1}^{N} 1_{w \in O_i}} \qquad (4)$$

where $n_{w,i}$ represents the number of times the word appears in the ith tweet, and NE_i represents the total number of words in the ith tweet. N represents the total number of tweets in the dataset, and O_i indicates the ith tweet. An example of TF-IDF
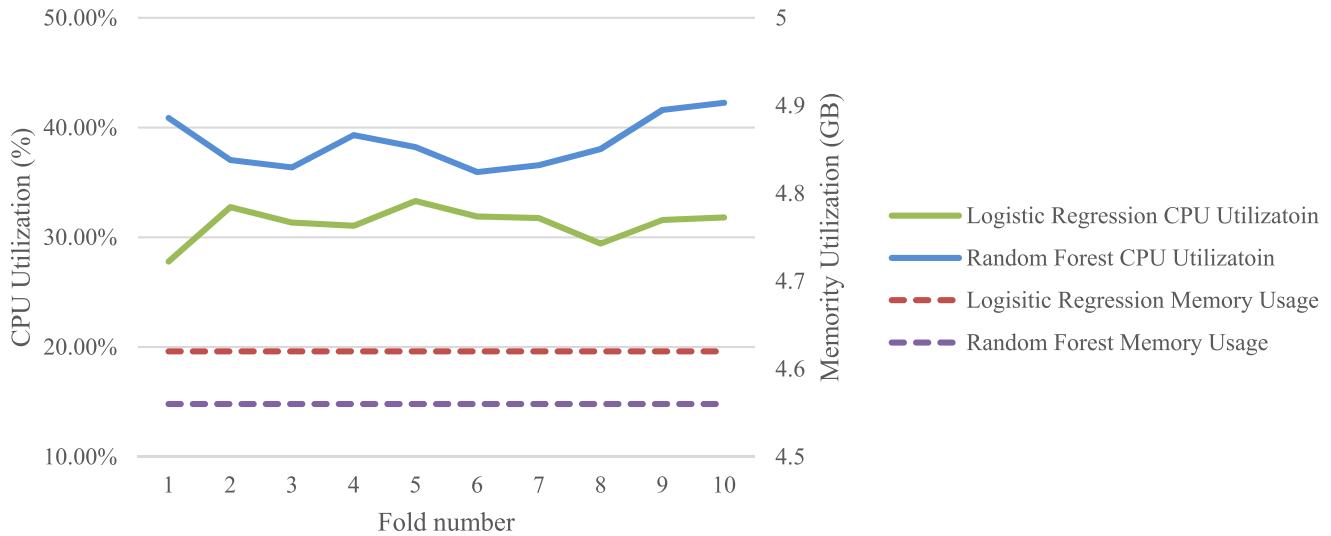
FIGURE 18. CPU and memory utilizations of Random Forest and Logistic Regression models during training.
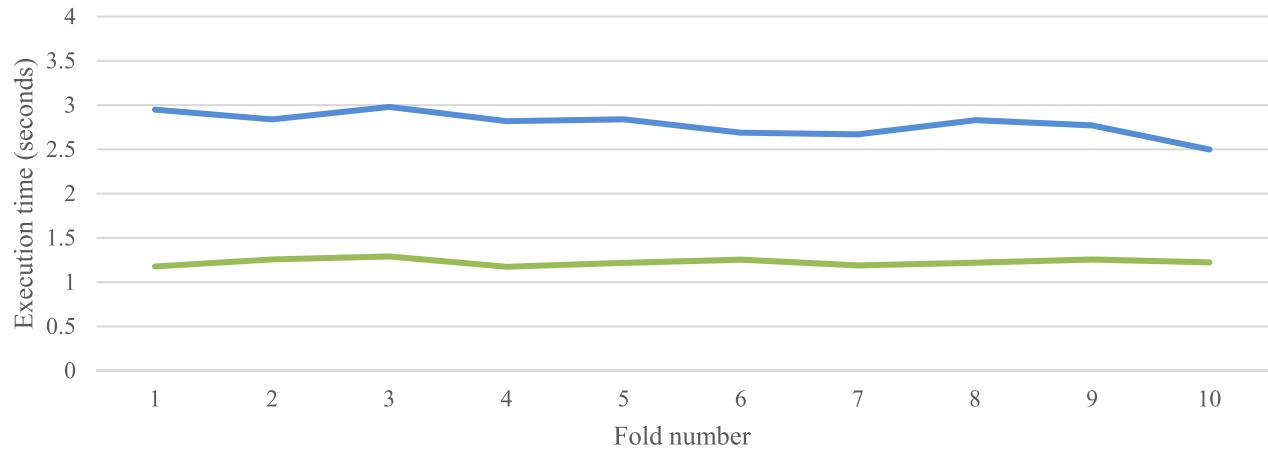

FIGURE 19. Execution time of Random Forest and Logistic Regression models during training.

is available in Appendix C. Appendix D shows an example of the feature extraction using concatenated BoW and TF-IDF.

Stage 6: Model building

We select Random Forest (RF) and Logistic Regression (LR) to evaluate our proposed framework, as these are the most used machine learning models from the literature as illustrated in Fig. 12.

C. RANDOM FOREST (RF)

RF, a supervised classification machine learning model based on a decision tree, builds an ensemble of decision trees by first selecting bootstrapped samples and random subset of features (words) from the dataset. It selects a feature that most cleanly splits the current dataset into parts that minimize impurity, facilitating class separation. It does so recursively until each of the final splits corresponds to only a single class [64]. At test time, an unseen tweet is passed through the tree, choosing paths according to its feature values that

minimize the impurity. Such predictions from multiple trees are collected, and the final emotion is predicted by a majority vote. This is done by counting the votes for each class that the individual decision trees have assigned to the tweet. The emotion that receives the highest number of votes, is ultimately chosen as the final predicted emotion for the unseen tweet, as seen in Equation (5).

$$E_T = majority \; vote \; \{E_n \, (T)\} \qquad (5)$$

where E_T is the emotion of a tweet and n is the number of decision trees.

D. LOGISTIC REGRESSION (LR)

LR is another supervised machine learning model which predicts one of two classes by (1) multiplying the feature vector representing a tweet with a learned weight vector and (2) passing the result through a Sigmoid function that predicts the likelihood of a tweet belonging to a certain emotion or

TABLE 5. Output of preprocessing steps for the UAE COVID-19 dataset.

Pre-processing step	Output
Remove duplicate observations	Remove 19,268 tweets.
Sample raw observation from the UAE COVID-19 dataset: **Get well soon. Great gesture from the playing 11 for their COVID19 affected teammates. @Alhilal_EN @TheAFCCL @Alhilal_FC https://t.co/Gm1DZePJdh**	
Remove usernames and links	Get well soon. Great gesture from the playing 11 for their COVID19 affected teammates.
Remove punctuation marks	Get well soon Great gesture from the playing for their COVID affected teammates
Tokenization	['Get', 'well', 'soon', 'Great', 'gesture', 'from', 'the', 'playing', 'for', 'their', COVID, 'affected', 'teammates']
Case folding	['get', 'well', 'soon', 'great', 'gesture', 'from', 'the', 'playing', 'for', 'their', 'covid', 'affected', 'teammates']
Remove stop words	['get', 'well', 'soon', 'great', 'gesture', 'playing', 'covid', 'affected', 'teammates']
Lemmatization	['get', 'well', 'soon', 'great', 'gesture', 'playing', 'covid', 'affected', 'teammate']

class on a scale from 0 (the tweet certainly does not express the emotion) to 1 (the tweet expresses the emotion with 100% certainty) [65]. In this case study, we will use LR as a multiclass model, using Equation (6).

$$P\,(Emotion) = \frac{e^{\beta_0 + \sum_{i=1}^{n} \beta_i T_i}}{1 + e^{\beta_0 + \sum_{i=1}^{n} \beta_i T_i}} \qquad (6)$$

where $P\,(Emotion)$ represents the probability of each emotion classification, T is the tweet, and β_0 and β_i are the regression coefficients representing the intercept and the slope respectively. The regression coefficient values are calculated using maximum likelihood estimation, as stated in Equation (7).

$$l\,(\beta_0 \ldots \beta_i) = \prod_{i,y_i=1} P(Emotions)$$
$$\times \prod_{i,y_i=0} (1 - P\,(Emotions)) \qquad (7)$$

E. EXPERIMENTAL ENVIRONMENT AND HYPERPARAMETER TUNING

To evaluate the performance of our framework, we use the COVID-19-vaccine-events-related dataset, which consists of 813 tweets and emotion labels, as shown in Figure 10. The experiments were run on Ubuntu 20.04 with 1024 GB RAM and 2 AMD EPYC 7552 48-Core, 96-Thread, 2.20 GHz Processors, using Python 3.9. Through Python, we used the following packages: Keras, NumPy, Pandas, Scikit-learn, TensorFlow, and Matplotlib for plotting the figures. As shown in Table 4, for the hyperparameter tuning, we used the optimal values obtained in the literature for each parameter of the algorithms under study. The optimal parameters used in

our experiments are calculated based on the GridSearchCV method.

Stage 7: Emotion prediction model evaluation

To develop the machine learning model, we split the dataset using 10-fold cross-validation. Moreover, we validate the model using the following evaluation metrics:

- Accuracy: It is the rate of correct classifications by the model. The accuracy score lies between 0 and 1, with 1 indicating that the model correctly classifies all the observations in the testing dataset.
- Precision: It is defined as the correct class classification rate out of all the class predictions. The precision value lies between 0 and 1, with 1 indicating that every observation classified in a class belongs to that class.
- Recall: It is defined as the correct class classification rate out of all the class observations. The recall value lies between 0 and 1, with 1 indicating that all the observations from a class are indeed classified as belonging to that class.
- F-measure: It is defined as the harmonic mean of precision and recall. The value of the F-measure lies between 0 and 1, with 1 indicating a perfect value of precision and recall.
- ROC Curve: It is a probabilistic curve representing the performance of a classification model at different classification thresholds. This is by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR).
- AUC: It measures the classifier's ability to differentiate between classes. AUC is calculated as the

TABLE 6. Feature extraction using Bag of Words (BoW).

Sample preprocessed observations
Observation 1: ['get', 'well', 'soon', 'great', 'gesture', 'playing', 'covid', 'affected', 'teammate']
Observation 2: ['second', 'shot', 'covid', 'vaccine']
BoW step 1
The vocabulary of unique words: 'get', 'well', 'soon', 'great', 'gesture', 'playing', 'covid', 'affected', 'teammate', 'second', 'shot', 'vaccine'
BoW step 2

	1 get	2 well	3 soon	4 great	5 gesture	6 playing	7 covid	8 affected	9 teammate	10 second	11 shot	12 vaccine
Observation 1	1	1	1	1	1	1	1	1	1	0	0	0
Observation 2	0	0	0	0	0	0	1	0	0	1	1	1

Features' vectors using BoW

The features' vector for observation 1 : [1 1 1 1 1 1 1 1 1 0 0 0]

The features' vector for observation 2: [0 0 0 0 0 0 1 0 0 1 1 1]

two-dimensional area under the ROC curve. Its value lies between 0 and 1, with 1 indicating that the classifier perfectly differentiates between the classes and 0 indicating that all the positive observations are classified as negative and all the negative ones as positive.

- Total Execution Time: It is defined as the total time the classifier takes for training (model building) and validation.
- CPU Usage: This measures the percentage of the CPU's capacity being utilized during the machine learning models' execution.
- Memory Usage: This measures the percentage of the RAM's capacity being utilized during the machine learning models' execution.

Fig. 13 shows the accuracy, F-measure, AUC, precision, and recall, of the RF and LR classifiers. It indicates that, in all cases, RF outperformed LR. This is because RF captures non-linear relationships between the words (features) and the corresponding emotion class [67]. In addition, our dataset is imbalanced as the distribution of the 9 emotion classes are inequal among the tweets. RF is more suitable than LR in handling imbalanced data. This is because it selects balanced samples of all predicted emotion classes at the time of building each decision tree. Consequently, the training dataset considered by RF is a balanced one [13], [68].

Finally, we present the RF and LR's ROC curves and the AUC values in Figs. 14 and 15 respectively. The ROC curve for the sadness emotion in both classifiers is equal to 1 as all tweets are classified correctly, bearing in mind the small number of tweets classified with such emotion. Additionally, the ROC curve for all emotions increased or remained the same for RF compared to LR, indicating RF detected more true values than LR for all emotions. We conclude that the RF is most suitable for classifying tweets' emotions. Finally, Figs. 16 and 17 illustrate the confusion matrix of RF and LR, respectively. The insights of the figures are in line with the ROC curves, as it is clear that there are more emotions correctly classified in RF than LR, which explains the performance illustrated in Fig. 13.

Machine Learning algorithms should be time efficient as time to analyze the emotions for events such as a pandemic is critical to develop a proactive prevention plan. Figs. 18 and 19 show the CPU and memory utilizations, and execution times respectively while training RF and LR for model development. Fig. 18 show that LR consumes less CPU than RF, even though both models consider all features. This is because RF builds 310 trees while developing the model, where each tree is considered as a model, unlike LR [13]. The figure also shows that both models have similar memory usage. However, RF takes longer time to run than LR, as shown in Fig. 19. This is because the time complexity for RF is $O(n \cdot f \cdot d \cdot t)$, where n refers to the number of trees, f refers to the number of features, d is the maximum depth of the trees,

TABLE 7. Feature extraction using Term Frequency – Inverse Document Frequency (TF-IDF).

Sample preprocessed observations
Observation 1: ['get', 'well', 'soon', 'great', 'gesture', 'playing', 'covid', 'affected', 'teammate']
Observation 2: ['second', 'shot', 'covid', 'vaccine']

Calculation of TF for the words in observations 1 and 2

	1 get	2 well	3 soon	4 great	5 gesture	6 playing	7 covid	8 affected	9 teammate	10 second	11 shot	12 vaccine
Observation 1	$TF_{1,1}$ $= \frac{1}{9}$	$TF_{2,1}$ $= \frac{1}{9}$	$TF_{3,1}$ $= \frac{1}{9}$	$TF_{4,1}$ $= \frac{1}{9}$	$TF_{5,1}$ $= \frac{1}{9}$	$TF_{6,1}$ $= \frac{1}{9}$	$TF_{7,1}$ $= \frac{1}{9}$	$TF_{8,1}$ $= \frac{1}{9}$	$TF_{9,1} = \frac{1}{9}$	$TF_{10,1}$ $= 0$	$TF_{11,1}$ $= 0$	$TF_{12,1}$ $= 0$
Observation 2	$TF_{1,2}$ $= 0$	$TF_{2,2}$ $= 0$	$TF_{3,2}$ $= 0$	$TF_{4,2}$ $= 0$	$TF_{4,2} 0$	$TF_{5,2}$ $= 0$	$TF_{6,2}$ $= \frac{1}{4}$	$TF_{7,2}$ $= 0$	$TF_{8,2} = 0$	$TF_{10,2}$ $= \frac{1}{4}$	$TF_{10,2}$ $= \frac{1}{4}$	$TF_{10,2}$ $= \frac{1}{4}$

Calculation of IDF for the vocabulary of words

1 get	2 well	3 soon	4 great	5 gesture	6 playing	7 covid	8 affected	9 teamamte	10 second	11 shot	12 vaccine
IDF_1 $= 0.3010$	IDF_2 $= 0.3010$	IDF_3 $= 0.3010$	IDF_4 $= 0.3010$	IDF_5 $= 0.3010$	IDF_6 $= 0.3010$	IDF_7 $= 0$	IDF_8 $= 0.3010$	IDF_9 $= 0.3010$	$IDF_{10} =$ 0.3010	$IDF_{11} =$ 0.3010	$IDF_{12} =$ 0.3010

Calculation of TF-IDF for the words in observations 1 and 2

	1 get	2 well	3 soon	4 great	5 gesture	6 playing	7 covid	8 affected	9 teamamte	10 second	11 shot	12 vaccine
Observation 1	TF $-IDF_{1,1}$ $= 0.033$	TF $-IDF_{2,1}$ $= 0.033$	TF $-IDF_{3,1}$ $= 0.033$	TF $-IDF_{4,1}$ $= 0.033$	TF $-IDF_{5,1}$ $= 0.033$	TF $-IDF_{6,1}$ $= 0.033$	TF $-IDF_{7,1}$ $= 0$	TF $-IDF_{8,1}$ $= 0.033$	TF $-IDF_{9,1}$ $= 0.033$	TF $-IDF_{10,1}$ $= 0$	TF $-IDF_{11,1}$ $= 0$	TF $-IDF_{12,1}$ $= 0$
Observation 2	TF $-IDF_{1,2}$ $= 0$	$TF_{2,2}$ $-IDF$ $= 0$	$TF_{3,2}$ $-IDF$ $= 0$	$TF_{4,2}$ $-IDF$ $= 0$	$TF_{5,2}$ $-IDF$ $= 0$	$TF_{6,2}$ $-IDF$ $= 0$	$TF_{7,2}$ $-IDF$ $= 0$	$TF_{8,2}$ $-IDF$ $= 0$	$TF_{9,2}$ $-IDF$ $= 0$	TF $-IDF_{10,2}$ $= 0.075$	TF $-IDF_{11,2}$ $= 0.075$	TF $-IDF_{12,2}$ $= 0.075$

Features' vectors using TF-IDF

The features vector for observation 1: [0.03 0.03 0.03 0.03 0.03 0 0.03 0.03 0 0 0]

The features vector for observation 2: [0 0 0 0 0 0 0 0 0 0.08 0.08 0.08]

and t is the number of tweets in the dataset; while the time complexity for LR $O(f \cdot t)$.

IV. DISCOUSION

A. PRINCIPAL FINDINGS

This study argues for a need for a proactive event-based framework that determines public emotions and tackles population-level emotions associated with events such as a pandemic and its related measurements. The effect that COVID-19 has on the human body is enormous [8], [9], [10], and its complications led to the death of over 6 million individuals globally [16]. However, the measures the policymakers take to contain the virus affect the individual's emotional state, whether testing positive for COVID-19 or not. In general, people who experience less positive and more negative emotions are more likely to become depressed

TABLE 8. Feature extraction using concatenated Bag of Words (BoW) and Term Frequency – Inverse Document Frequency (TF-IDF).

Sample preprocessed observations

Observation 1: ['get', 'well', 'soon', 'great', 'gesture', 'playing', 'covid', 'affected', 'teammate']

Observation 2: ['second', 'shot', 'covid', 'vaccine']

Concatenation of BoW and TF-IFD for the words in observations 1 and 2

	BoW												TF-IDF											
	1 get	2 well	3 soon	4 great	5 gesture	6 playing	7 covid	8 affected	9 teammate	10 second	11 shot	12 vaccine	1 get	2 well	3 soon	4 great	5 gesture	6 playing	7 covid	8 affected	9 teammate	10 second	11 shot	12 vaccine
Observation 1	1	1	1	1	1	1	1	1	1	0	0	0	0.03	0.03	0.03	0.03	0.03	0.03	0	0.03	0.03	0	0	0
Observation 2	0	0	0	0	0	0	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0.08	0.08	0.08

Features' vectors using TF-IDF

Feature vector for observation 1 : [1 1 1 1 1 1 1 1 1 0 0 0 0.03 0.03 0.03 0.03 0.03 0 0.03 0.03 0 0 0]

Feature vector for observation 2: [0 0 0 0 0 0 1 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0.08 0.08 0.08]

and have suicidal thoughts. Similarly, there is a correlation between the heightened experience of anger and a rise in suicide attempts [66]. Furthermore, a study in the United States reported that suicidal behavior among high school students was significantly positively correlated with sadness [67]. During the COVID-19 pandemic, researchers found that lockdowns generated fear emotions, causing increased levels of anxiety, depression, and other mental health disorders [22], [23], [68]. Significant events in people's lives affect their emotions. Therefore, it becomes critical to detect population emotions, particularly the negative ones, to avoid mental health issues and their consequences. Monitoring emotions in response to particular events on a societal level is essential. In particular, following the introduction of policies and legislation, it can aid policymakers in responding effectively and avoiding the surge of levels of distress which may lead to political disorders or violence.

Consequently, researchers predict the sentiments associated with the pandemic using different machine-learning models [25], [30], [31], [32], [33], [34], [35], [36], [37], [69]. However, rather than fine-grained and psychologically founded emotions as considered in our paper, previous work predominantly focused on coarser-grained positive, negative, and neutral sentiments, and lack a systematic approach to emotion detection in response to a particular event. Therefore,

in this paper, we propose an event-based machine-learning framework that distinguishes three basic sentiments (positive, negative, or neutral) and Plutchik's eight emotions (joy, fear, anger, anticipation, sadness, surprise, disgust, and trust). It can shed light on the population-level emotional state through a methodological process that enables the emotional detection of social media posts, in response to the occurrence of a particular event.

Based on our manual annotation, we found that positive emotions dominated the UAE-COVID-19-vaccine-related tweets overall (a positive sentiment, and trust emotion). This suggests a level of trust among the social media users in the UAE's government and policymakers in placing preventive and precautionary measures to contain the virus. In addition, the UAE Twitter population expressed more negative sentiment before the vaccine commencement (September 14, 2020), although fear slightly increased and trust slightly decreased after vaccine commencement, although some of these emotions could be related to movement restrictions rather than vaccines per se. Moreover, we trained and evaluated Random Forest and Logistic Regression to automatically predict emotions for tweets. Based on a variety of evaluation metrics, we consistently found that the random forest outperformed the logistic regression, achieving 95.5% accuracy, 95.8% precision, 95.4% recall, 95.6% F-measure, and 98.1% AUC, with almost the same memory utilization during

training. However, RF is 2.3 times slower and consumes around 7% more CPU than LR.

B. LESSONS LEARNED

- *Annotation tools*: Based on our results, we suggest data scientists use more than one automatic data annotation tools such as TextBlob, VADER, and NRCLex and to verify the labels produced by these tools manually. This is because as found in our experiments after manual annotation, these tools produce contradictory results. This can be life-threatening if negative tweets with sad and depressed emotions are labeled as positive tweets.
- *Event-based emotion analysis*: Researchers and government organizations studying public emotions during a pandemic should consider the intervention of different events within the framework. This will aid in developing timely tailored pandemic prevention and public health strategies.

V. CONCLUSION

The COVID-19 pandemic created a sense of fear and other negative emotions. Additionally, the events associated with the pandemic and the increasing number of infections and deaths led to negative emotions affecting mental health. Several works in the literature studied the emotions associated with COVID-19 through machine learning. However, this paper is the first to propose a machine learning-based framework to detect emotions based on the occurrence of specific events and uses the UAE as a case study and the COVID-19 vaccine as an event. In addition, we detect 11 emotion classes related to introducing precautionary measures or events. We implemented the proposed framework using the ML widely used machine learning algorithms Random Forest (RF) and Logistic Regression (LR). Our experimental results show that RF outperforms LR.

There are a few limitations in this paper, including the fact that what individuals share on social media is not necessarily what they feel; the text is not a direct signal of emotions that would warrant judging the mental health state on an individual level. We argue, however, that social media expressions are a valuable indicator for a population-level response. Moreover, the Twitter population is not representative of the general UAE population, therefore, countermeasures should not be based solely on tweets. However, our framework presents a generic systematic approach to detecting emotions and is not specific to a particular social media type or a sample of the population. It can be applied to a population at the national level, considering all social media types, for better insights into population emotions to provide signals to policymakers following the introduction of measures or legislation.

APPENDIX A

Table 5 shows the output for the pre-processing steps on a sample observation taken from UAE COVID-19 dataset.

APPENDIX B

As shown in Table 6, two pre-processed tweets were selected randomly from the dataset. These tweets are converted into feature vectors using BoW. In step 1, as shown in the table, the list of unique words from both observations is created. Next, in step 2, a vector for each tweet is created, having the frequency of term frequency of the unique words in that tweet. These feature vectors for each tweet will then be used for training the learning algorithm.

However, with BoW, important words occurring less frequently can be neglected. As a result, the technique gives more weight to the most common words, which can lead to inaccurate results in sentiment analysis.

APPENDIX C

Table 7 shows the development of the feature vectors for two pre-processed sampled tweets using TF-IDF. For each word, the TF is computed, and then the IDF. The features' vectors for each tweet are then represented using the TF-IDF value of each word in each tweet.

APPENDIX D

Table 8 shows an example of concatenating BoW and TF-IDF as part of the feature extraction stage.

REFERENCES

[1] Aristotle, *Nicomachean Ethics*. Indianapolis, Indiana: Hackett Publishing Company, 2019.

[2] R. Plutchik and H. Kellerman, *Emotion: Theory, Research, and Experience*, vol. 1. Academic, 1980.

[3] T. Dalgleish and M. Power, *Handbook of Cognition and Emotion*. Hoboken, NJ, USA: Wiley, 2000.

[4] C. E. Izard, "Emotion theory and research: Highlights, unanswered questions, and emerging issues," *Annu. Rev. Psychol.*, vol. 60, no. 1, pp. 1–25, Jan. 2009, doi: 10.1146/annurev.psych.60.110707.163539.

[5] L. Ismail, H. Materwala, A. P. Karduck, and A. Adem, "Requirements of health data management systems for biomedical care and research: Scoping review," *J. Med. Internet Res.*, vol. 22, no. 7, Jul. 2020, Art. no. e17508, doi: 10.2196/17508.

[6] L. Cruz-Villalobos and L. Barret. (2001). *The Nature of Emotions*. [Online]. Available: http://www.jstor.org/page/info/about/policies/terms.jsp

[7] World Health Organization. *Timeline: WHO's COVID-19 Response*. Accessed: Jan. 9, 2023. [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/interactive-timeline

[8] S. D. Unudurthi, P. Luthra, R. J. C. Bose, J. R. McCarthy, and M. I. Kontaridis, "Cardiac inflammation in COVID-19: Lessons from heart failure," *Life Sci.*, vol. 260, Nov. 2020, Art. no. 118482, doi: 10.1016/j.lfs.2020.118482.

[9] M. D. Shah, A. S. Sumeh, M. Sheraz, M. S. Kavitha, B. A. V. Maran, and K. F. Rodrigues, "A mini-review on the impact of COVID 19 on vital organs," *Biomed. Pharmacotherapy*, vol. 143, Nov. 2021, Art. no. 112158, doi: 10.1016/j.biopha.2021.112158.

[10] A. Abdel-Moneim and A. Hosni, "Insights into the possible impact of COVID-19 on the endocrine system," *Arch. Physiol. Biochem.*, vol. 129, no. 4, pp. 998–1006, Mar. 2021, doi: 10.1080/13813455.2021.1890131.

[11] L. Ismail and H. Materwala, "From conception to deployment: Intelligent stroke prediction framework using machine learning and performance evaluation," in *Proc. IEEE Int. Conf. Omni-Layer Intell. Syst. (COINS)*, Aug. 2022, pp. 1–7, doi: 10.1109/COINS54846.2022.9854961.

[12] A. Hennebelle, L. Ismail, H. Materwala, J. Al Kaabi, P. Ranjan, and R. Janardhanan, "Secure and privacy-preserving automated machine learning operations into end-to-end integrated IoT-edge-artificial intelligence-blockchain monitoring system for diabetes mellitus prediction," 2022, *arXiv:2211.07643*.

[13] L. Ismail, H. Materwala, M. Tayefi, P. Ngo, and A. P. Karduck, "Type 2 diabetes with artificial intelligence machine learning: Methods and evaluation," *Arch. Comput. Methods Eng.*, vol. 29, no. 1, pp. 313–333, Jan. 2022, doi: 10.1007/s11831-021-09582-x.

[14] L. Ismail, H. Materwala, and J. Al Kaabi, "Association of risk factors with type 2 diabetes: A systematic review," *Comput. Struct. Biotechnol. J.*, vol. 19, pp. 1759–1785, 2021, doi: 10.1016/j.csbj.2021.03.003.

[15] L. Ismail and H. Materwala, "Comparative analysis of machine learning models for diabetes mellitus type 2 prediction," in *Proc. Int. Conf. Comput. Sci. Comput. Intell. (CSCI)*, Dec. 2020, pp. 527–533, doi: 10.1109/CSCI51800.2020.00095.

[16] National Emergency Crisis and Disasters Management Authority. *Recent News | UAE Coronavirus (COVID-19) Updates*. Accessed: Jan. 9, 2023. [Online]. Available: https://covid19.ncema.gov.ae/en

[17] C. Hou, J. Chen, Y. Zhou, L. Hua, J. Yuan, S. He, Y. Guo, S. Zhang, Q. Jia, C. Zhao, J. Zhang, G. Xu, and E. Jia, "The effectiveness of quarantine of Wuhan City against the corona virus disease 2019 (COVID-19): A well-mixed SEIR model analysis," *J. Med. Virol.*, vol. 92, no. 7, pp. 841–848, Jul. 2020, doi: 10.1002/jmv.25827.

[18] L. Ismail, H. Materwala, Y. A. Hammadi, F. Firouzi, G. Khan, and S. R. B. Azzuhri, "Automated artificial intelligence-enabled proactive preparedness real-time system for accurate prediction of COVID-19 infections—Performance evaluation," *Frontiers Med.*, vol. 9, pp. 1–13, Aug. 2022, doi: 10.3389/fmed.2022.871885.

[19] A. D. Kaye, C. N. Okeagu, A. D. Pham, R. A. Silva, J. J. Hurley, B. L. Arron, N. Sarfraz, H. N. Lee, G. E. Ghali, J. W. Gamble, H. Liu, R. D. Urman, and E. M. Cornett, "Economic impact of COVID-19 pandemic on healthcare facilities and systems: International perspectives," *Best Pract. Res. Clin. Anaesthesiol.*, vol. 35, no. 3, pp. 293–306, Oct. 2021, doi: 10.1016/j.bpa.2020.11.009.

[20] M. Škare, D. R. Soriano, and M. Porada-Rochon, "Impact of COVID-19 on the travel and tourism industry," *Technol. Forecasting Social Change*, vol. 163, Feb. 2021, Art. no. 120469, doi: 10.1016/j.techfore.2020.120469.

[21] B. Pfefferbaum and C. S. North, "Mental health and the COVID-19 pandemic," *New England J. Med.*, vol. 383, no. 6, pp. 510–512, Aug. 2020, doi: 10.1056/NEJMp2008017.

[22] M. Wathelet, S. Duhem, G. Vaiva, T. Baubet, E. Habran, E. Veerapa, C. Debien, S. Molenda, M. Horn, P. Grandgenèvre, C.-E. Notredame, and F. D'Hondt, "Factors associated with mental health disorders among university students in France confined during the COVID-19 pandemic," *JAMA Netw. Open*, vol. 3, no. 10, Oct. 2020, Art. no. e2025591, doi: 10.1001/jamanetworkopen.2020.25591.

[23] M. É. Czeisler, R. I. Lane, E. Petrosky, J. F. Wiley, A. Christensen, R. Njai, M. D. Weaver, R. Robbins, E. R. Facer-Childs, L. K. Barger, C. A. Czeisler, M. E. Howard, and S. M. W. Rajaratnam, "Mental health, substance use, and suicidal ideation during the COVID-19 pandemic—United States, June 24–30, 2020," *MMWR. Morbidity Mortality Weekly Rep.*, vol. 69, no. 32, pp. 1049–1057, Aug. 2020, doi: 10.15585/mmwr.mm6932a1.

[24] D. Kim, C. J. Chung, and K. Eom, "Measuring online public opinion for decision making: Application of deep learning on political context," *Sustainability*, vol. 14, no. 7, p. 4113, Mar. 2022, doi: 10.3390/su14074113.

[25] C. Singh, T. Imam, S. Wibowo, and S. Grandhi, "A deep learning approach for sentiment analysis of COVID-19 reviews," *Appl. Sci.*, vol. 12, no. 8, p. 3709, Apr. 2022, doi: 10.3390/app12083709.

[26] S. M. Mohammad and P. D. Turney, "Crowdsourcing a word–emotion association lexicon," *Comput. Intell.*, vol. 29, no. 3, pp. 436–465, Aug. 2013.

[27] R. Plutchik, *The Emotions*. New York, NY, USA: Random House, 1962.

[28] R. Plutchik, "A general psychoevolutionary theory of emotion," in *Theories of Emotion*. Amsterdam, The Netherlands: Elsevier, 1980, pp. 3–33, doi: 10.1016/B978-0-12-558701-3.50007-7.

[29] R. Plutchik, *The Psychology and Biology of Emotion*. New York, NY, USA: Harper Collins, 1990.

[30] K. Garcia and L. Berton, "Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA," *Appl. Soft Comput.*, vol. 101, Mar. 2021, Art. no. 107057, doi: 10.1016/j.asoc.2020.107057.

[31] R. Chandra and A. Krishna, "COVID-19 sentiment analysis via deep learning during the rise of novel cases," *PLoS ONE*, vol. 16, no. 8, Aug. 2021, Art. no. e0255615, doi: 10.1371/journal.pone.0255615.

[32] 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*, vol. 16, no. 2, Feb. 2021, Art. no. e0245909, doi: 10.1371/journal.pone.0245909.

[33] U. Naseem, I. Razzak, M. Khushi, P. W. Eklund, and J. Kim, "COVID-Senti: A large-scale benchmark Twitter data set for COVID-19 sentiment analysis," *IEEE Trans. Computat. Social Syst.*, vol. 8, no. 4, pp. 1003–1015, Aug. 2021, doi: 10.1109/TCSS.2021.3051189.

[34] 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," *Appl. Soft Comput.*, vol. 97, Dec. 2020, Art. no. 106754, doi: 10.1016/j.asoc.2020.106754.

[35] S. Agrawal, S. K. Jain, S. Sharma, and A. Khatri, "COVID-19 public opinion: A Twitter healthcare data processing using machine learning methodologies," *Int. J. Environ. Res. Public Health*, vol. 20, no. 1, p. 432, Dec. 2022, doi: 10.3390/ijerph20010432.

[36] C. Villavicencio, J. J. Macrohon, X. A. Inbaraj, J.-H. Jeng, and J.-G. Hsieh, "Twitter sentiment analysis towards COVID-19 vaccines in the Philippines using Naïve Bayes," *Information*, vol. 12, no. 5, p. 204, May 2021, doi: 10.3390/info12050204.

[37] A. Shahzad, B. Zafar, N. Ali, U. Jamil, A. J. Alghadhban, M. Assam, N. A. Ghamry, and E. T. Eldin, "COVID-19 vaccines related user's response categorization using machine learning techniques," *Computation*, vol. 10, no. 8, p. 141, Aug. 2022, doi: 10.3390/computation10080141.

[38] E. H. Hovy, "What are sentiment, affect, and emotion? Applying the methodology of Michael Zock to sentiment analysis," in *Language Production, Cognition, and the Lexicon*. Cham, Switzerland: Springer, 2015, pp. 13–24, doi: 10.1007/978-3-319-08043-7_2.

[39] H. H. Do, P. Prasad, A. Maag, and A. Alsadoon, "Deep learning for aspect-based sentiment analysis: A comparative review," *Expert Syst. Appl.*, vol. 118, pp. 272–299, Mar. 2019, doi: 10.1016/j.eswa.2018.10.003.

[40] A. M. Schoene, L. Bojanic, M.-Q. Nghiem, I. M. Hunt, and S. Ananiadou, "Classifying suicide-related content and emotions on Twitter using graph convolutional neural networks," *IEEE Trans. Affect. Comput.*, vol. 14, no. 3, pp. 1791–1802, Jul./Sep. 2023, doi: 10.1109/TAFFC.2022.3221683.

[41] S. J. Lee, J. Lim, L. Paas, and H. S. Ahn, "Transformer transfer learning emotion detection model: Synchronizing socially agreed and self-reported emotions in big data," *Neural Comput. Appl.*, vol. 35, no. 15, pp. 10945–10956, May 2023, doi: 10.1007/s00521-023-08276-8.

[42] *TextBlob: Simplified Text Processing*. Accessed: Jun. 23, 2023. [Online]. Available: https://textblob.readthedocs.io/en/dev/

[43] *vaderSentiment*. Accessed: Jun. 23, 2023. [Online]. Available: https://pypi.org/project/vaderSentiment/

[44] A. Glenn, P. LaCasse, and B. Cox, "Emotion classification of Indonesian tweets using bidirectional LSTM," *Neural Comput. Appl.*, vol. 35, no. 13, pp. 9567–9578, May 2023, doi: 10.1007/s00521-022-08186-1.

[45] M. Krommyda, A. Rigos, K. Bouklas, and A. Amditis, "An experimental analysis of data annotation methodologies for emotion detection in short text posted on social media," *Informatics*, vol. 8, no. 1, p. 19, Mar. 2021, doi: 10.3390/informatics8010019.

[46] L. Ismail, N. Shahin, H. Materwala, A. Hennebelle, and L. Frermann. *ML-SMEmot Dataset*. Accessed: May 1, 2023. [Online]. Available: https://github.com/INDUCE-Lab/ML-SMEmot

[47] L. Ismail, B. Mills, and A. Hennebelle, "A formal model of dynamic resource allocation in grid computing environment," in *Proc. 9th ACIS Int. Conf. Softw. Eng., Artif. Intell., Netw., Parallel/Distrib. Comput.*, 2008, pp. 685–693.

[48] L. Ismail, H. Materwala, and A. Hennebelle, "A scoping review of integrated blockchain-cloud (BcC) architecture for healthcare: Applications, challenges and solutions," *Sensors*, vol. 21, no. 11, p. 3753, May 2021.

[49] L. Ismail, "Dynamic resource allocation mechanisms for grid computing environment," in *Proc. 3rd Int. Conf. Testbeds Res. Infrastruct. Develop. Netw. Communities*, 2007, pp. 1–5, doi: 10.1109/TRIDENTCOM.2007.4444737.

[50] J. J. Webster and C. Kit, "Tokenization as the initial phase in NLP," in *Proc. 14th Conf. Comput. Linguistics*, 1992, pp. 1106–1110.

[51] M. Toman, R. Tesar, and K. Jezek, "Influence of word normalization on text classification," in *Proc. INSCIT*, vol. 4, 2006, pp. 354–358.

[52] D. Jurafsky and J. H. Martin, *Speech and Language Processing An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Upper Saddle River, NJ, USA: Prentice-Hall, 2020.

[53] T. Fushiki, "Estimation of prediction error by using K-fold cross-validation," *Statist. Comput.*, vol. 21, no. 2, pp. 137–146, Apr. 2011, doi: 10.1007/s11222-009-9153-8.

[54] L. Ismail and H. Materwala, "IDMPF: Intelligent diabetes mellitus prediction framework using machine learning," *Appl. Comput. Informat.*, vol. 2021, pp. 1–10, Jun. 2021, doi: 10.1108/ACI-10-2020-0094.

[55] L. Ismail, H. Materwala, T. Znati, S. Turaev, and M. A. B. Khan, "Tailoring time series models for forecasting coronavirus spread: Case studies of 187 countries," *Comput. Struct. Biotechnol. J.*, vol. 18, pp. 2972–3206, Jan. 2020, doi: 10.1016/j.csbj.2020.09.015.

[56] L. Ismail, H. Materwala, and A. Hennebelle, "Forecasting COVID-19 infections in Gulf cooperation council (GCC) countries using machine learning," in *Proc. 13th Int. Conf. Comput. Model. Simul.*, Jun. 2021, pp. 231–236, doi: 10.1145/3474963.3475844.

[57] L. Ismail, S. Alhmoudi, and S. Alkatheri, "Time series forecasting of COVID-19 infections in United Arab Emirates using ARIMA," in *Proc. Int. Conf. Comput. Sci. Comput. Intell. (CSCI)*, Dec. 2020, pp. 801–806, doi: 10.1109/CSCI51800.2020.00150.

[58] United Arab Emirates Government. *Fact Sheet*. Accessed: Mar. 8, 2023. [Online]. Available: https://u.ae/en/about-the-uae/fact-sheet

[59] P. Siemund, A. Al-Issa, and J. R. E. Leimgruber, "Multilingualism and the role of English in the United Arab Emirates," *World Englishes*, vol. 40, no. 2, pp. 191–204, Jun. 2021, doi: 10.1111/weng.12507.

[60] National Emergency Crisis and Disaster Management Authority. *UAE Authorises Emergency Use of COVID19 Vaccine for Members of First Line of Defence*. Accessed: Mar. 8, 2023. [Online]. Available: https://covid19.ncema.gov.ae/en/News/Details/1462

[61] *NRCLex*. Accessed: Jun. 23, 2023. [Online]. Available: https://pypi.org/project/NRCLex/#description

[62] H. Lane, C. Howard, and H. M. Hapke. (2018). *Natural Language Processing in Action Understanding, Analyzing, and Generating text With Python*. [Online]. Available: http://www.manning.com

[63] Y. Zhang, R. Jin, and Z.-H. Zhou, "Understanding bag-of-words model: A statistical framework," *Int. J. Mach. Learn. Cybern.*, vol. 1, nos. 1–4, pp. 43–52, Dec. 2010, doi: 10.1007/s13042-010-0001-0.

[64] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Inf. Process. Manage.*, vol. 24, no. 5, pp. 513–523, Aug. 1988, doi: 10.1016/0306-4573(88)90021-0.

[65] B. Pfahringer, "Random model trees: An effective and scalable regression method," Univ. Waikato, Hamilton, New Zealand, Tech. Rep. 03/2010, 2010.

[66] D. W. Homser Jr., S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*. Wiley, 2013.

[67] P. Zhao, X. Su, T. Ge, and J. Fan, "Propensity score and proximity matching using random forest," *Contemp. Clin. Trials*, vol. 47, pp. 85–92, Mar. 2016, doi: 10.1016/j.cct.2015.12.012.

[68] H. Zhao, X. Chen, T. Nguyen, J. Z. Huang, G. Williams, and H. Chen, "Stratified over-sampling bagging method for random forests on imbalanced data," in *Proc. Pacific-Asia Workshop Intell. Secur. Inform.*, 2016, pp. 63–72, doi: 10.1007/978-3-319-31863-9_5.

[69] L. Seidlitz, Y. Conwell, P. Duberstein, C. Cox, and D. Denning, "Emotion traits in older suicide attempters and non-attempters," *J. Affect Disord*, vol. 66, nos. 2–3, pp. 123–131, Oct. 2001, doi: 10.1016/S0165-0327(00)00300-1.

[70] D. G. Trimble and A. Chandran, "Associations between sad feelings and suicide behaviors in the 2019 youth risk behavior survey: A call for action," *Frontiers Pediatrics*, vol. 9, pp. 1–11, Sep. 2021, doi: 10.3389/fped.2021.694819.

[71] N. K. Fofana, F. Latif, S. Sarfraz, M. F. Bashir, and B. Komal, "Fear and agony of the pandemic leading to stress and mental illness: An emerging crisis in the novel coronavirus (COVID-19) outbreak," *Psychiatry Res.*, vol. 291, Sep. 2020, Art. no. 113230, doi: 10.1016/j.psychres.2020.113230.

[72] S. Yu, D. Eisenman, and Z. Han, "Temporal dynamics of public emotions during the COVID-19 pandemic at the epicenter of the outbreak: Sentiment analysis of Weibo posts from Wuhan," *J. Med. Internet Res.*, vol. 23, no. 3, Mar. 2021, Art. no. e27078, doi: 10.2196/27078.

LEILA ISMAIL (Member, IEEE) received the joint Ph.D. degree (Hons.) from the Grenoble Institute of Technology and French Institute for Research in Computer Science and Technology (INRIA), France. She is the Founding Director of the Intelligent Distributed Computing and Systems (INDUCE) Laboratory, United Arab Emirates University (UAEU), an Associate Professor with the Department of Computer Science and Software Engineering, and a Visiting Associate Professor at Cloud Computing and Distributed Systems (CLOUDS) Laboratory, School of Computing and Information Systems, Faculty of Engineering and Information Technology, The University of Melbourne, Australia. She has vast industrial and academic experience with the Sun Microsystems Research and Development Center, France, on the design and implementation of real-time and highly available distributed systems, and participated in the deposit of a U.S. patent. She served as a Lecturer with the Grenoble Institute of Technology and an Adjunct Professor with the Digital Ecosystems and Business Intelligence Institute, Curtin University, Australia. She has been very active in creating smart and efficient digital ecosystems responding to nowadays emergency needs for better living in our dynamic global habitat, introducing artificial intelligence, machine learning, deep learning, and blockchain approaches for different smart city applications domains, such as energy savings, transportation systems, and healthcare. She is active in international collaborations and publishes her research results in prestigious journals and international conferences. She served as an Associate Editor for the *International Journal of Parallel, Emergent and Distributed Systems* for several years. She served as the chair, the co-chair, and the track chair for many IEEE and ACM international conferences, including being the General Co-Chair for IEEE DEST 2009, and the General Chair, the Technical Program Chair, and the Organizing Committee Chair for the 11th International Conference on Innovations in Information Technology 2015 (IIT'15) for which she got the support of the IEEE Computer Society (HQs, USA) technical sponsorship. She is an Editor of the *Information Innovation Technology in Smart Cities* (Nature Springer, 2018). She has been invited as a keynote speaker to several conferences, including the Women in Data Science International Conference (WiDS 2021), organized by Stanford University, USA. She is a recipient of several awards and appreciation certificates.

NADA SHAHIN received the master's degree in information technology management from the University of Wollongong, in 2020. She is currently pursuing the Ph.D. degree with the Intelligent Distributed Computing and Systems (INDUCE) Laboratory, College of Information Technology, United Arab Emirates University (UAEU), United Arab Emirates. She was awarded the Basic Fellowship Program from UAEU plus graduate research assistantship to support her studies.

HUNED MATERWALA received the master's degree in control systems from the Department of Electrical and Electronics Engineering, Amity University, Uttar Pradesh, India, in 2016. He is currently pursuing the Ph.D. degree with the Intelligent Distributed Computing and Systems (INDUCE) Laboratory, College of Information Technology, United Arab Emirates University (UAEU), United Arab Emirates. He has been a Research Assistant with the INDUCE Laboratory and awarded a Ph.D. Scholarship from UAEU to support his studies. His research interests include the Internet of Things, edge, and cloud computing, energy efficiency and management, and smart healthcare for sustainable smart cities applications.

LEA FRERMANN is currently a Senior Lecturer (an Assistant Professor) and a DECRA Fellow with The University of Melbourne. Her research aims to understand how humans learn about and represent complex information and to enable models to do the same in fair and robust ways. To do so, she combines natural language processing and machine learning with the cognitive and social sciences. Recent projects include models of meaning change; of common sense knowledge in human and computational representations of language; and automatic story understanding in both fiction (books or movies) and the real world (media narratives on issues like climate change).

● ● ●

ALAIN HENNEBELLE received the master's degree in computer science from Ecole des Mines de St-Étienne and University of St-Etienne, France, and the micro-master's degree in statistics and data sciences from the Massachusetts Institute of Technology (MIT), USA. He is currently a Teaching Professional with the School of Computing and Information Systems, The University of Melbourne and Deakin University, Melbourne, Australia. He is also a Chief Architect with extensive experience in artificial intelligence, machine learning, data science, software engineering, and development operations (DevOps), working in international companies, such as Sun Microsystems (Oracle), Bull (France) in cooperation with IBM. His research interests include machine learning algorithms and statistics, cloud computing, and smart city applications.