Journal of Biomedical Informatics 137 (2023) 104258

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Methodological Review

Textual emotion detection in health: Advances and applications

Alieh Hajizadeh Saffar [a], Tiffany Katharine Mann [b], Bahadorreza Ofoghi [b,*]

[a] *Ferdowsi University of Mashhad, Mashhad, Iran*
[b] *School of Information Technology, Deakin University, Burwood Victoria 3125, Australia*

ARTICLE INFO

Keywords:
Natural language processing
Machine learning
Health applications
Medical applications
Physical health
Mental health
Textual emotion detection

ABSTRACT

Textual Emotion Detection (TED) is a rapidly growing area in Natural Language Processing (NLP) that aims to detect emotions expressed through text. In this paper, we provide a review of the latest research and development in TED as applied in health and medicine. We focus on medical and non-medical data types, use cases, and methods where TED has been integral in supporting decision-making. The application of NLP technologies in health, and particularly TED, requires high confidence that these technologies and technology-aided treatment will first, do no harm. Therefore, this review also aims to assess the accuracy of TED systems and provide an update on the state of the technology. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines were used in this review. With a specific focus on the identification of different human emotions in text, the more general sentiment analysis studies that only recognize the polarity of text were excluded. A total of 66 papers met the inclusion criteria. This review found that TED in health and medicine is mainly used in the detection of depression, suicidal ideation, and the mental status of patients with asthma, Alzheimer's disease, cancer, and diabetes with major data sources of social media, healthcare services, and counseling centers. Approximately, 44% of the research in the domain is related to COVID-19, investigating the public health response to vaccinations and the emotional response of the public. In most cases, deep learning-based NLP techniques were found to be preferred over other methods due to their superior performance. Developing methods for implementing and evaluating dimensional emotional models, resolving annotation challenges by utilizing health-related lexicons, and using deep learning techniques for multi-faceted and real-time applications were found to be among the main avenues for further development of TED applications in health.

1. Introduction

Textual Emotion Detection (TED) is an important application of Natural Language Processing (NLP) facilitating automated analysis and detection of the primary emotion expressed in text. While a more general NLP application of sentiment analysis usually focuses on the coarse-grained classification of text into positive, neutral, or negative polarity [1], fine-grained analysis using TED generally uses scales of emotional labels, such as the six basic emotions introduced by Ekman [2]: sadness, anger, surprise, fear, happiness, and disgust.

NLP techniques and processes have been utilized widely in the health and biomedical domains [3,4] and TED in particular has had significant impacts on several medical applications. Understanding moods and emotions regarding health events [5,6], early detection and early public health intervention [7,8], studying the impact of negative news headlines on individual emotional well-being [9], and mental health counseling support chatbots [10,11] are among the recent applications of TED in health.

While the number of TED resources and applications grows, an important concern arises. Is the application of TED technologies to health and biomedical problems acceptably accurate and effective? These technologies in medical domains to provide or support treatment and assessment for humans require high confidence and assurance that such technology-backed applications will first, do no harm. A thorough review of the research in this area allows an assessment of the most precise NLP and TED approaches and provides an update on the state of the technology. Several review studies on methods, challenges, and advances in textual emotion recognition [12–14] and sentiment analysis [15,16] have been presented in recent years. There are also studies on the role of NLP and text analytics in the broader healthcare space [17], such as social media-based surveillance systems [18] and more particularly, in mental health [19,20]. In addition, the work in [21] investigates sentiment analysis in health and well-being. From several perspectives, however, a review study on the application of emotion recognition and analysis in health and medical applications

* Corresponding author.
E-mail address: b.ofoghi@deakin.edu.au (B. Ofoghi).

https://doi.org/10.1016/j.jbi.2022.104258
Received 20 June 2022; Received in revised form 24 November 2022; Accepted 27 November 2022
Available online 14 December 2022

is still lacking. There remain several under-studied research questions in the domain which this article endeavors to address, and therefore its main contributions fall within the following areas:

- Health applications, including various healthcare areas, of physical well-being, mental health, and improved medical systems.
- Data types and sources of data in TED applications including clinical and non-clinical text.
- Techniques, models, and evaluation metrics, which are used to detect and evaluate fine-grained emotions in text, in health and medical applications.
- Other resources, that although are mostly open-domain and not specific to health, have implications for TED and thus impact the effectiveness of TED systems in health.
- Challenges, limitations, and future trends of current health-related TED techniques, that will shed light on the future of TED in the domain.

The rest of this article is organized as follows. Section 2 describes the strategy of article search, selection, and data synthesis. In Section 3, after summarizing review statistics and a brief introduction to the emotion analysis models, the reviewed articles are categorized according to the different criteria outlined above (emotion models, data sources, TED applications and methods in health, and TED evaluation). A summary of the main findings and analytic discussions regarding current limitations and future trends of TED in health are given in Section 4 and, Section 5 concludes the article.

2. Methodology

This work follows PRISMA 2020 [22], a 27-item checklist and four-phase flow diagram to improve transparency in systematic reviews. The PRISMA checklist covers the minimum set of items required throughout a systematic review. It aims to help authors improve reporting of sections including title, abstract, introduction, methods, results, discussion, and funding in an integrated and standardized manner. A flow diagram details the number of identified records and excluded articles.

Commencing December 2021, a search for relevant articles was conducted by the authors using Scopus, PubMed, IEEE, ACM publications, ScienceDirect and CINAHL databases, and The Cochrane Library. Scopus was selected as the primary tool and the search query was determined based on the specific domain of the study, TED in health. Articles including three categories of keywords “natural language processing”, “emotion detection”, and “health” in their title, abstract, and set of keywords were of interest. Keywords in each category were identified through studying previous surveys on NLP [17,19,20], emotion detection [12–15], and sentiment analysis in health and well-being [21]. Review of these studies helped in the selection of focused health-related keywords; however, in order to exclude sentiment analysis studies and focus on the applications of fine-grained emotion identification, the word “sentiment” was excluded from the search query. The final list of articles was restricted to peer-reviewed journal articles and conference papers using the search query presented in Fig. 1.

The same strategy was followed in the search for articles in all data sources; however, in the case of IEEE, only abstracts were checked for their inclusion of the search keywords. Health-related keywords were removed when searching PubMed, CINAHL, and The Cochrane Library, as these are dedicated resources for the health domain and contextual health-related keywords were unnecessary. Due to restrictions on logical operators in ScienceDirect, several rounds of searches were performed using different terms and a limited number of keywords for each domain.

Duplicate articles were removed from the collection, and a number of articles were excluded after an initial review of titles and abstracts if they did not satisfy one or more of the inclusion criteria. Inclusion criteria are defined in Table 1.

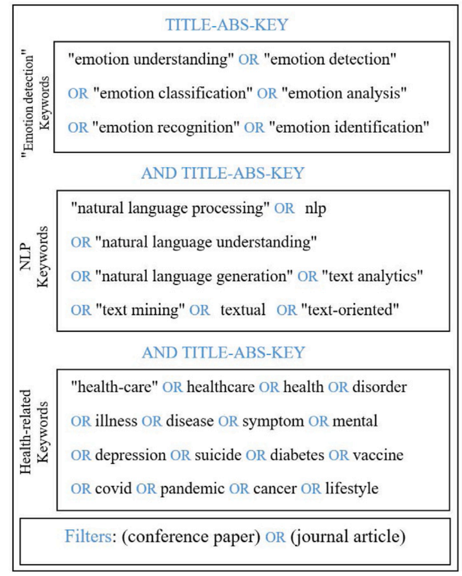


Fig. 1. Search query used to find relevant journal articles and conference papers on TED in health and medicine.

Two researchers were independently responsible for reading and collecting necessary, targeted data from the articles. Nearly 70% of the articles were studied by the first author while the remaining were reviewed by the second author. Some articles were removed from consideration at this stage when it was found that they in fact failed to meet the inclusion criteria.

A number of articles were reviewed by the third author to finalize the inclusion decision where they were ambiguous or lacked a clear methodology. In addition, after reviewing the references of the initial articles, a number of new papers were added to the list for further review if they met the specific aforementioned inclusion criteria. The targeted information extracted from the articles included: the main objective and novelty, health domain and application, data resources, data language, and data collection methods, emotion models, TED approaches and measures of performance, limitations, and challenges, future trends, and other potential developments. Information from each article was collected by the authors. The main findings were synthesized by the authors and the data were categorized for analysis with respect to each of the above items.

To assess the risk of bias in systematic reviews, the ROBIS [23] tool examines potential concerns in four domains: eligibility criteria, study selection, data extraction, and synthesis, by asking relevant questions and providing appropriate criteria. In this work, eligibility and study selection criteria are met using a clear set of pre-defined research questions and objectives, the relevant eligibility criteria were selected and are detailed in Table 1, a structured search strategy using targeted keywords and restrictions on various databases was utilized and references were checked to ensure no relevant articles were overlooked. Screening more than 50% of the articles by at least two authors also reduced the risk of bias in article selection.

Regarding data extraction, more than 30% of the articles were reviewed by both the first and the second authors. In this case, an inter-rater agreement of 0.70 (moderate agreement) based on Cohen’s Kappa statistic [24] was calculated using the confusion matrix shown in Table 2. The minimum sample size S required to test the null hypothesis with Kappa [25,26] ($\kappa \in [0, 0.2]$ for no to very low agreement [24]) versus the alternative hypothesis (κ=0.7), assuming a significance level of 0.05 and a power of 80%, is $S \in [11, 24]$ while our sample size (reviewed by two authors) was 29. This ensures a similar, moderate agreement rate on the entire set of articles reviewed. In addition, the

Table 1
Four main inclusion criteria for journal articles and conference papers.

Item	Criterion
1.	The subject of the article, its application, and/or the data used should be relevant to health and medicine.
2.	The article should detect the main, fine-grained categories of emotions from text, not only the polarity of sentiments.
3.	The article should include rigorous, relevant evaluations and discussion with reference to well-described datasets. A few articles without numeric evaluations of TED systems were included on the basis of useful applications discussed.
4.	The body of the article should be written in English.

Table 2
Confusion matrix for inclusion of the reviewed articles.

		Author 1		Row marginal
		Include	Exclude	
Author 2	Include	21	3	24
	Exclude	0	5	5
	Column marginal	21	8	29

research team included members with varied backgrounds in machine learning, health, and NLP to analyze the selected articles from different perspectives. Synthesis of findings is unlikely to produce biased results due to the set of inclusion and quality criteria for articles. All selected articles were included in the reported findings in accordance with the research questions. Further, the findings include a wide variety of data sources, applications, and methodologies.

3. Results

A total of 198 papers were identified, 92 from Scopus, 32 from PubMed, 23 from IEEE Explore, 8 from ACM publications, 4 from CINAHL, and 39 from ScienceDirect. After removing duplicate and inaccessible papers, the abstracts of 139 articles were reviewed for compliance with the selection criteria (Table 1). The search query included general words such as health, emotion, and text, resulting in retrieving some irrelevant articles, 58 of which were identified and excluded. For instance, the abstracts of many articles expressed that emotion detection has applications in health; however, the data and/or applications mentioned in the article were not related to health and medicine. While assessing the full texts, 29 articles were identified which did not meet inclusion criteria, and 14 articles were added to the list after analyzing the references of the remaining articles. Finally, 66 articles were selected for information extraction and data analysis. A flowchart of the search and filtering strategy utilized in this systematic review is presented in Fig. 2.

There was no publication year limit in the search for articles yet the oldest article in the final collection dates back to 2010. Although several studies exist on TED prior to this date, no health-related articles were found using the keywords described. Among the final papers, 15% were published between 2010 and 2016, most of which relate to the second track of the 2011 i2b2 NLP Challenge, of emotion classification in suicide notes. 18% of the articles were published between 2017 and 2019, with the remaining published in 2020 and 2021. This demonstrates that the use of TED in health applications has grown exponentially since 2019, largely due to the COVID-19 pandemic, indeed 44% of the articles reviewed were related to COVID-19.

3.1. Emotion models and theories in medical applications

With roots in psychology, two main emotion models are used to represent emotions in scientific terms [15]: *categorical* models and *dimensional* models. In categorical models, emotions are categorized into distinct and independent classes while dimensional models consider emotional states to be related and interdependent. Emotions in a dimensional space are represented with relative scores [27,28].

The most widely used categorical emotional models were introduced by Paul Ekman [2] and Robert Plutchik [29]. Ekman proposed a list of human emotions *happiness, fear, sadness, surprise, disgust,* and *anger*. Plutchik added *anticipation* and *trust* to Ekman's list. A much wider representation of emotion was proposed by Orthony, Clore, and Collins (OCC) [30] which contains several other categories in addition to Ekman's basic emotions.

Fig. 3 shows three structures of emotion spaces for dimensional models. The wheel of emotions introduced by Plutchik [29] (Fig. 3(c)) and Russell's circumplex [31] model (Fig. 3(a)) describe 2D models while Jin and Wang's 3D model [32] considers "power" as a third emotional dimension (Fig. 3(b)). Furthermore, Parrott [33] proposed a hierarchical structure that includes primary, secondary, and tertiary emotions.

Almost 45% of the papers studied made use of the 6-scale and 8-scale emotion models introduced by Ekman and Plutchik [6,34–36]. In some cases, one or two emotion labels were ignored or added to these well-known lists. For example, in [37,38], categories such as *sarcasm*, *criticism*, and *news-related* were added to the emotion list to account for the specific syndromic surveillance context of Ebola. In [39], *skepticism*, a non-basic emotion in the context of COVID-19, was added to Plutchik's emotion list. There are numerous research studies that select more diverse and varied emotion labels. The selection of such emotion categories is affected by the capabilities of the TED tools of open-source services (e.g., FastText in [40]) or labeled datasets used by [27,35,41,42].

Among the articles analyzed, few utilized dimensional emotion models. In [27], long and short texts in the Real World Worry Dataset (RWWD) were annotated by a set of emotions with a 9-scale system. Data and fine-grained annotations which captured the emotional responses of UK residents to COVID-19, collected by a crowd-sourcing platform, made it possible to predict emotions through the estimation of continuous scores. In [28], a vector of strength per emotion class was assigned to each sentence in a chatbot conversation. This vector was represented as $P_i = (x_{anger}, x_{disgust}, x_{joy}, x_{surprise}, x_{fear}, x_{contempt}, x_{sadness})$, where P_i is the ith user statement within the conversation with the chatbot and $x_{emotion}$ represents the normalized strength of the *emotion* from the three categories of low, medium, and high.

3.2. Data sources

Various sources were accessed within the domain of health and medicine to gather TED research data. These data were divided into two categories: *clinical* data and *non-clinical* data, as summarized in Table 3. Clinical data includes data collected from patient–physician interactions, usually written or created to receive medical guidance, and data collected from medical sensors. The majority of non-clinical data included social media posts (e.g., tweets) and labeled data collected from social media based on pre-defined, specific criteria. Due to frequent social, political, and individual concerns discussed on social

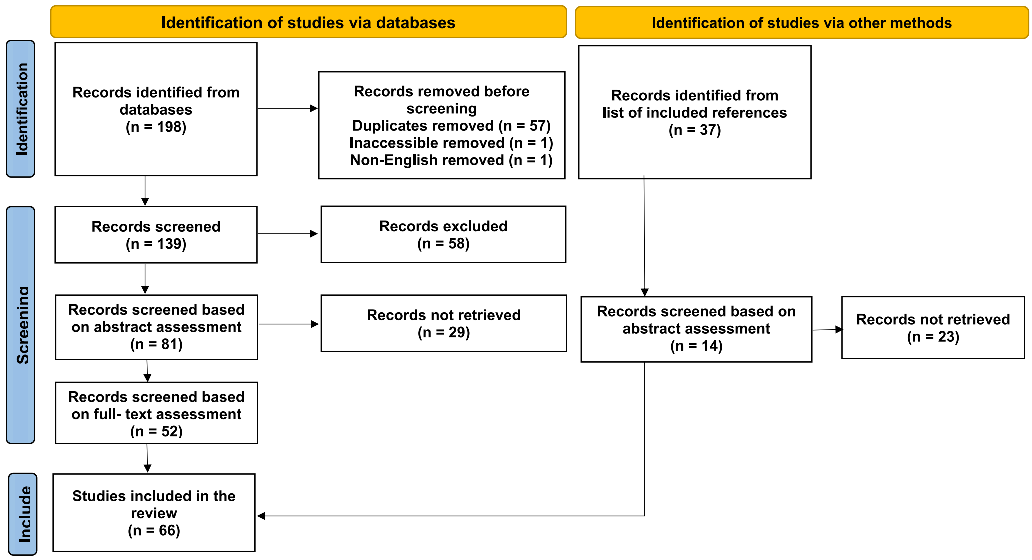
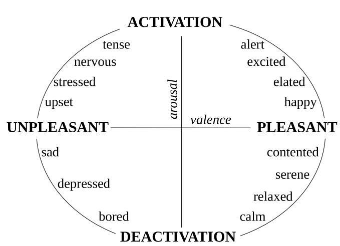
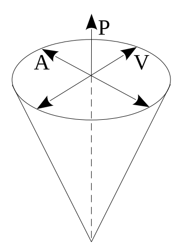


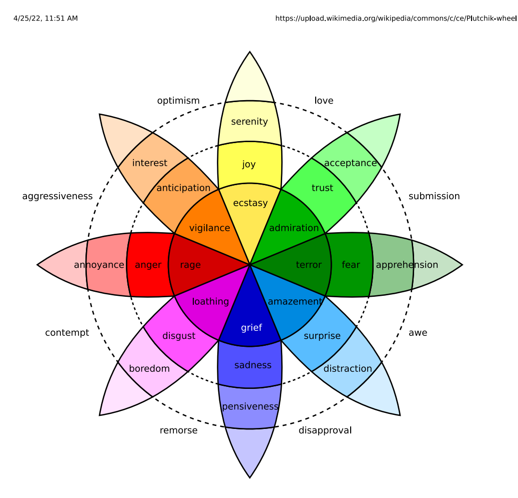
Fig. 2. Flow diagram of search strategy for reviewed articles.



(a) Russell's circumplex model [31]



(b) Jin and Wang's 3D emotional model adapted from [32], P-Power, V-Valence, and A-Arousal



(c) Plutchik's wheel of emotions [29]

Fig. 3. Different dimensional models of emotion taken from [14].

media, these networks are useful platforms for measuring the public health status within the community. Most TED data sources were from Twitter and its Chinese counterpart platform Weibo.

To access social media posts, most researchers utilized Twitter and Weibo's application programming interfaces (APIs) and tools, in combination with scripting libraries such as Tweepy.[1] Twarc,[2] a command line tool [40], TWINT[3] an advanced Twitter scraping tool [41], and RTweet[4] package, utilized in [36], are the other tools for this purpose. In some cases, APIs have been tailored to access tweets with specific keywords and topics, such as COVID-19-related posts in India [5]. Other examples of COVID-19 datasets from Twitter include SenWave [43], Semeval-2018 Affect in Tweets [44], Covid Stance [45], and Twitter chatter [46]. Table 3 summarizes the data collected for health and medicine-related TED research. Greater detail can be found in Appendix A: Table A.6.

In some studies, datasets that are not necessarily health-related were used before the main TED data were processed [34,50,67,74,78,95]. The main goals of using these datasets included:

- Pre-training models for transfer learning: Using transfer learning, [50,67] pre-trained deep learning models with one task, to help fine-tune parameters for other tasks, achieving the same or better performance with higher efficiency and less labeled data. Chatbots proposed in [74,78] were trained with general domain emotion datasets such as ISEAR[5] to show better performance when facing conversations.
- Evaluating models: Evaluation of the predictive power of a proposed model in [34] and assessing the effectiveness of sentence embedding in [95].

A list of auxiliary datasets used for TED method development and training is provided in Appendix B: Table B.7.

1 https://www.tweepy.org/
2 https://github.com/DocNow/twarc
3 https://github.com/twintproject/twint
4 https://github.com/ropensci/rtweet

5 The International Survey on Emotion Antecedents and Reactions.

Table 3
Data sources for TED in health and medicine.

Clinical data	
Online medical healthcare platforms	Webmd[a], Healthtap[b] [47]
Online therapy	Horyzon [48], PsychPark[c] [49]
Clinical interviews	DAIC-WOZ[d] [50]
Medical sensors	Converted text from patient audio data [51]
Non-clinical data	
Social media	Twitter [5–8,34,36,37,40,41,52–65], Twitter-based datasets [41,57,66–69], Reddit [7,39,70], Weibo [8,35,42,68,71], Facebook, YouTube, Instagram [53,56,60,63], Official press websites, news articles, internet forums and blogs [38,56,60]
Chatbot conversations	[10,11,28,72–79]
Online Health Communities (OHC)	[35,80–84]
Global English news sources	[9]
Crowdsourcing and interview	Suicide notes [85–92], COVID-19 Real World Worry Dataset [27], Collection of conversation-like text [93]

[a] https://www.webmd.com/.
[b] https://www.healthtap.com/.
[c] http://www.psychpark.org.
[d] Distress Analysis Interview Corpus Wizard-of-Oz [94].

In selecting such TED datasets, three key factors were found to be important to researchers: (i) large datasets, easily accessible with open-source features, e.g., emotion in text, GoEmotion, and ISEAR, (ii) the domain-specificity of the datasets, e.g., [34,95] and (iii) the emotional labels included in the datasets, e.g., a unified dataset was used in [67].

Some studies combined datasets to fine-tune their model parameters [50,78], to create greater compatibility with the domain [95], and to balance data label distributions [66]. To address data imbalance, several other solutions were used including; oversampling a special class [50], specific selection of data in different classes [7], using one binary classifier per emotion without focusing on feature design and hyper-parameter tuning [86], and weighting different classes during model training [66,95].

3.3. Applications of TED in health

Our analysis found that health-related applications of TED are classified into three major categories. One used TED for a variety of analyses on social media and to measure the effects of a medical event on society [7,54,59]. These works enable a deeper analysis of management practices as well as macro government decision-making in public health. In the second category, TED plays an important role in the development of individual or group healthcare platforms such as chatbots [28,70,78] and useful educational applications in the medical domain [75]. The third category includes online healthcare communities that used TED-based analyses to plan and deliver their services [83,84]. The full list of previous studies that cover these three categories is given in Tables 4 and C.8. The following sections include a more detailed analysis of these various areas and the medical applications of TED.

3.3.1. Medical events in social media

Health organizations rely on data to understand and prioritize public health risks, supporting decision-making, funding allocation, and government-level response. While appropriate public health management is paramount, understanding the population's response to the current climate can be crucial to success. A generally fearful public will result in different messaging and education to a trusting public. Tracking how this changes over time can inform public education and track progress [59]. The emotion of the public regarding a number of health-related events has been analyzed using TED, including; the impact of the COVID-19 pandemic on mental health, influenza and influenza-like illnesses, Legionnaires' disease, and Ebola [34,37,55,70]. Table 4 presents the various analyses used in different health contexts to inform macro-health management decisions. One subject of analysis in each of these examples is COVID-19. With the onset of the pandemic, researchers examined the emotional impact of social distancing, lockdown, and isolation on the emotional response of communities over time. They found social media platforms to be excellent tools for analyzing and interpreting people's concerns. With social media analysis, a number of studies have focused on specific regions with geo-tagged tweets and posts [58], trending hashtags and keywords including India [69], Spain [56], and China [42,68] or languages other than English such as Arabic [80], Roman Urdu [63], Dutch, German, Italian, and Swedish [39].

3.3.2. Healthcare supporting tools and platforms

Health-related chatbots and mobile applications with conversational user interfaces [10,11,72,73], such as Wysa,[6] have mostly been designed to improve the mental health of their users by identifying mental illnesses and suggesting preventive treatment. Developing self-assessment tools [50], Extracting mental disorders [47,81], Virtual Standardized Patient (VSP) [75], and Internet of Health (IoH) Frameworks [51] are other examples of TED applications for the development, improvement, and analysis of healthcare platforms and chatbots. A detailed list of TED applications in healthcare platforms is provided in Appendix C: Table C.8. Numerous researchers utilized healthcare platform data to further develop technical aspects of emotion detection and improve the effectiveness of mental healthcare systems, with a focus on emotional characteristics, intensities, and transitions [81] and using suicide notes [85–92].

3.3.3. Service improvements in health communities

Numerous studies focused on analyzing patients' emotions to improve services provided by health communities or online forums. Examples include promoting patient-centric care in oculoplastic surgery [84], fluctuating emotions in prostate cancer patients from diagnosis through to recovery [83], the impact of public holidays on emotional states in breast and lung cancer patients [80], and classification of Facebook emotions of the diabetes community [82]. A complete list of goals of TED in online health associations is presented in Appendix B: Table C.8.

Table 4
Main objectives of emotion analysis in social media for different health domains.

Main goal	Medical event/health domain
Real-time monitoring of people's mood/emotion by web portals	COVID-19 [5], New York Legionnaires' Disease Outbreak [6]
Studying the emotional impact of medical events and analyzing people's emotional changes and mental health status over time	COVID-19 [36,39,40,42,52–54,56–58,61,64,65,68,69,93], Influenza [34,55]
Early detection/intervention in public health for better management	COVID-19 [7,8], Depression [41,63,71], Ebola [37]
Detecting adverse drug reactions	Pharmacovigilance [62]
Analysis of risk communication in a medical event	COVID-19 [60]
Understanding and improving public perception of a specific aspect of a medical condition/event	COVID-19 vaccination [59], Autism Spectrum Disorder [35]
Studying the impact of negative news headlines on emotional well-being	COVID-19 [9]
Fake news detection	COVID-19 [67]

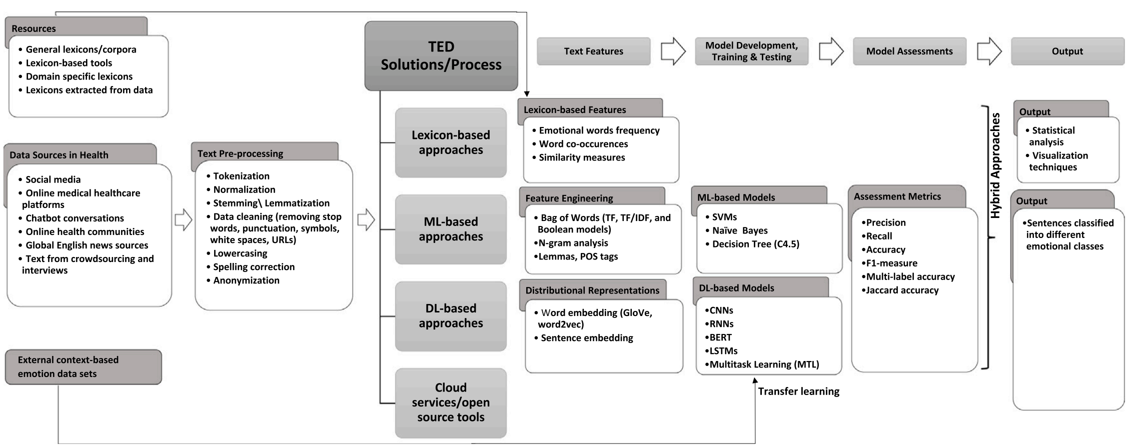


Fig. 4. TED data, solutions, and processes as applied in health and medicine.

3.4. TED methods in health and medicine

Text-based emotion recognition is mostly considered a supervised machine learning classification problem, to assign one or more emotion categories to a sentence or selection of text [14], including most previous work relating to TED in health. TED can also be defined as a problem of estimating different emotional scores for sentences [27], however, this is a less common approach taken by researchers in the domain. Table 5 presents different methods used for emotion identification. The process that takes place in TED for these types of methods can be seen in Fig. 4. The first step, common to most NLP tasks, is pre-processing textual data. Tokenization, lemmatization, removing stop words, and spelling correction are the most common data pre-processing operations. Cleaned and normalized text then goes through feature extraction, model training and development, and system evaluation. During feature extraction, the text is broken into sentences and words (through tokenization) and the tokens are converted into understandable numeric vectors for the machine. The TED model (of any type described below) will use such input vectors to learn mapping relationships between input text and output emotion labels or scores.

Lexicon-based approaches have been the most basic and traditional methods to distinguish between different types of emotions within text, utilizing keyword-based and corpus-based approaches [14]. In keyword-based approaches, the emotion score for every possible emotion within the text is calculated according to linguistic and statistical key features and rules [15] with the help of lexicons such as WordNet-Affect [96]. WordNet-Affect, an extended form of WordNet [97], contains affective words annotated with emotion labels [15]. Such lexicon-based approaches make use of categorical dictionaries with a limited number of categories. The performance of these techniques is affected by the ambiguity of keywords and lack of linguistic information and domain-specific words as traditional lexicons have limited information and flexibility [13]. To solve the problem of domain-specificity, corpus-based approaches have been developed with superior performance but poor generalizability through the use of corpora that contain random samples of text within particular domains. This approach is implemented by rule-based algorithms with NLP parsing techniques [15].

Numerous lexicons and dictionaries are used in lexicon-based methods as a source of emotion search words or keywords. Based on the data summarized in Appendix E: Table E.10, the Canadian National

[6] https://www.wysa.io/

Table 5
Methods used for TED in health and medicine.

TED methods	Type and references
Lexicon-based approaches	Keyword-based [5,9,10,35,36,39,42,54,58,59,75], Corpus-based [61,69]
Machine learning models	SVMs [49,55,63,73,85,86,91,92], GSOM[a] [64,81], Regression [27,34,90], Naïve Bayes [6,49,51,55,63], Maximum Entropy [87], C4.5 [49], KNN [63], Random Forest [63,65], SDMs[b]+Likelihood-based Learning [41]
Deep learning models	CNN-based [8,74], RNN-based [28,74], BERT-based [7,66–68,78], LSTM-based [11,50,53,57,67,70–72], HAN [74]
Hybrid models	ML+Lexicon-based model+CRF[c] [37,38,62,88,89], ML+DL [83], DL combinations [11,47,72,77,80,95]
Cloud models/ open-source frameworks and services	IBM's Watson [56,60,84], Indico API[d] [82], DialogFlow[e] [76,79], Facebook's Fasttext [40], Syuzhet[f] [52]

[a]Growing Self-Organising Maps.
[b]Smoothed Dirichlet Multinomial distribution.
[c]Conditional Random Field.
[d]https://indico.io/docs/emotion.
[e]https://cloud.google.com/dialogflow.
[f]https://github.com/mjockers/syuzhet.

Research Council (NRC) Word-Emotion Association Lexicon[7] [98,99] has been the most popular and widely used emotion lexicon. Some researchers constructed lexicons according to the problem requirements. For example, a hand-crafted lexicon of negation terms was constructed in [89], using a simple dictionary look-up method, to detect negation signals. The authors found that incorporating negation detection into the identification of emotion expressions is necessary because negation words (e.g., no, not, never) and phrases (e.g., no longer, do not) can intensify the emotion strength or shift the emotion polarity. Lexicons were utilized to extend emotion datasets with emotion synonyms [34], word clusters [86], and psychiatric labels [47]. Lexicon features were also used to train machine learning models [6,55,85,92], which form another category of text-based emotion classification. These methods are capable of learning new tasks without being specifically programmed for the new task, by dividing the entire dataset into two parts: (i) the training dataset for training model parameters and hyperparameters, and (ii) the testing dataset to understand how effective the model will be on new unseen data or tasks [15]. According to Table 5, Naïve Bayes, Support Vector Machines (SVMs), decision trees, and regression models were among the techniques used in health and medicine applications for emotion recognition. Bag of Words (BOWs), n-grams, lemmas, and POS tags are usually used as the input features for these models.

Recently, deep learning networks as a subdomain within machine learning, have proven their superiority in textual emotion recognition among other NLP tasks [15]. Most current deep learning models utilize word or sentence embeddings as input features which are transformations of words or sentences into continuous real-valued vectors. Such embeddings capture the semantic relatedness of words and sentences that have a similar meaning and are usually learned using neural networks or matrix factorization techniques. Word2Vec [100], Global Vector (GloVe) [101], FastText [102], and ELMO [103] are among the most widely used techniques for learning embedding representations.

Hybrid models combine approaches with the aim of overcoming limitations and shortcomings of various individual methods [15]. They are mostly a combination of machine learning models and lexicon-based methods or hybrid structures of deep learning networks (see Table 5). The need for combining machine learning algorithms is seen because each of these models performs well on specific emotional labels due to the complexity of emotion expressions and ambiguity inherent in natural language and where some labels have insufficient training instances.

Adding lexicon-based features to machine learning classifiers also has an incremental effect on the performance of the machine learning TED systems as seen in the combination of methods developed in [37,38,80,88,95].

Previous works have focused more on the application than on the method, making use of open-source services or cloud models for efficiency. These services and tools provide a package of machine learning methods. IBM's Watson [104], utilized in [56,60,84], is a cognitive computing platform that offers a wide range of services including Discovery, Knowledge Studio, Language Translator, Natural Language Classifier/Understanding, and Personality Insights with high precision [60]. DialogFlow,[8] a google Natural Language Understanding platform, is used for building chatbots with the help of rule-based methods and machine learning algorithms, to generate appropriate chatbot responses to user inputs [76,79]. Indico[9] is an API for text and image analysis [82] and Facebook's FastText is a library for text representations and classification [40]. The full list of the systems, platforms, and studies that have made use of these platforms is summarized in the final section of Table 5.

3.5. Evaluation metrics and performance analysis

Statistical analysis is used to measure the effectiveness of lexicon-based models regarding the frequency of emotional words or, lexicon-based calculation of the emotional scores of sentences. Scores are then presented using visualization techniques such as histograms or radar plots [10,35,42,54]. To automatically distinguish emotions from text through machine learning-based models, however, model performance on training and test datasets using standard evaluation metrics, on different models [49,55,63,81] or on a model with different feature settings [49,55,86,87,90] must be measured. The most widely used performance evaluation criteria for TED include precision, recall, F1 score, and accuracy. To compute the end-to-end performance of a TED system with multiple emotion labels, metrics should be extended to deal with a multi-label classification problem and to account for any imbalanced data. Micro and macro-averaged F1 scores, multi-label accuracy, and Jaccard accuracy are among the most widely used criteria in such systems [66]. Multi-label accuracy is computed according to Eq. (1), where D is the number of testing samples, m is the number of labels,

[7] http://saifmohammad.com/WebPages/ NRC-Emotion-Lexicon.htm.

[8] https://cloud.google.com/dialogflow

[9] https://indico.io/docs/emotion

y_{ij} is the jth ground truth label of sample i, and $\hat{y}_{ij}$ is its corresponding predicted label.

$$MultiLabelAcc. = \frac{1}{D \times m} \sum_{i=1}^{D} \sum_{j=1}^{m} \alpha(\hat{y}_{ij} == y_{ij}) \tag{1}$$

It is not always possible to use the evaluation results and criteria reported in the literature to compare the performance across TED applications. This is particularly true when the problems/applications, data distributions, and feature engineering methods have obvious differences across different studies or systems. However, on the basis of machine learning-based measures alone, the effectiveness of SVMs has proven superior among several machine learning approaches including Naïve Bayes and C4.5 [49,55]. Similarly, authors in [8,68] have shown deep learning methods are more accurate than traditional machine learning methods. In deep learning methods, two factors significantly increase the effectiveness of TED: (i) unsupervised learning algorithms to develop word or sentence embeddings where resembling semantics are represented by similar vectors [105], and (ii) transfer learning techniques to make use of the knowledge contained in similar contexts and apply it in new domains (e.g., [50,67]).

The effectiveness of TED systems has been analyzed with a combination of lexicon-based features and machine learning models in [38,88]. From these studies alone, a reliable conclusion cannot be reached on the increased performance of TED systems by combining deep learning methods with each other or with auxiliary features. Neither study compared the performance of non-hybrid or bare methods (without lexicon-based features) with that of combined models on the same problem. Work by [80] however, makes such evaluation possible by analyzing performance outcomes of a hybrid deep learning model (CNNs+LSTMs) combined with lexicon-based features, compared with the individual CNN and LSTM models without the use of any external lexicon. Given no other similar work was found to compare the performance of hybrid models with those of individual base models without the use of lexicon-based features, we believe there remains room for further research in this direction to assess the effect of auxiliary and external features on deep learning-based TED systems.

In terms of evaluating multidimensional emotion models, Mean Absolute Error and coefficient of determination (r^2) were used to evaluate the emotion scores estimated by a linear regression model in [27]. However, no articles reviewed evaluated the overall TED results of multidimensional models and no comparison of different emotional models was found.

3.6. Summary statistics

This review demonstrates that a large proportion of previous studies and TED systems have used categorical emotion models. Classification results of the current TED systems are therefore easily distinguishable, especially where a clear separation of diagnosis and relevant support is required. Some articles implemented multi-label classification [7,47,66,68] and there were only a few studies that modeled emotion intensity and variation (e.g., [28,64,81]).

As shown in Fig. 5(a), the most commonly used data for TED in health are spontaneously generated narratives on public platforms expressing public concerns. Identifying the emotions of the public can be considered as public health monitoring or investigation of public response regarding specific medical events including epidemics. Clinical data, mainly provided by online counseling systems, also contribute a small share towards TED research.

Fig. 5(b) demonstrates the most prominent TED application in health is to analyze text composed by social media users. Fig. 5(d) shows how TED has been used less in patients with diseases other than pandemics (such as COVID-19 and Influenza) and psychological disorders.

With a focus on models, Fig. 5(c) shows that deep learning models are among the most widely used, coinciding with deep learning models being the most accurate solutions for TED in health on several occasions [8,68].

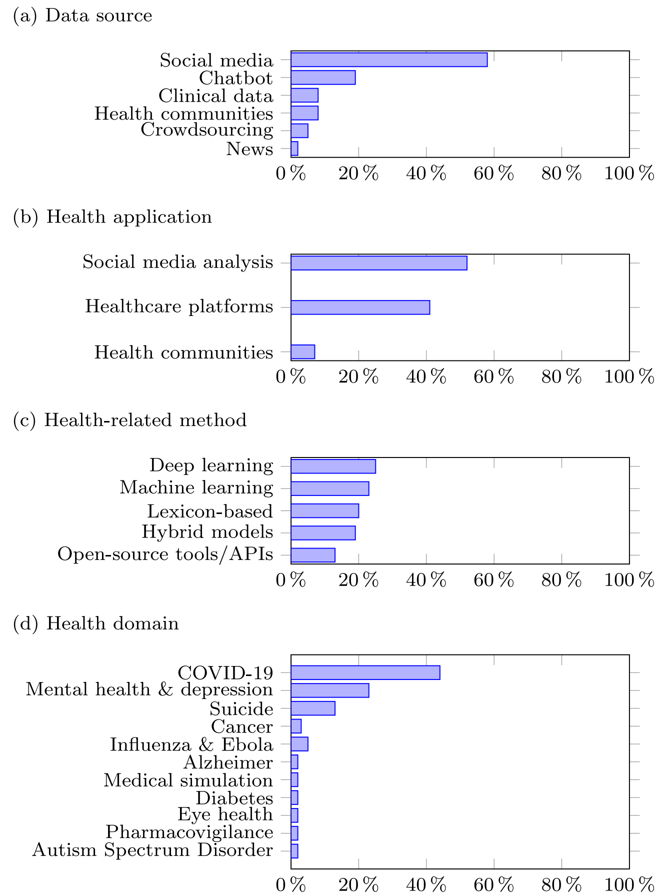


Fig. 5. Categorical statistics on data, applications, health domains, and TED methods as applied in health and medicine. These statistics are derived from the reviewed articles.

4. Discussion

To review the current state of the research, the development and application of TED in the health domain can be analyzed from several angles. This analysis has practical implications for TED researchers and those who intend to further develop TED applications in health.

More effective TED in health: Annotation of sentences with emotion labels is dependent on the preference or perception of the annotator or on a selected set of emotions. This shows that accurately identifying real human emotions is a complex task, more sensitive in health as this involves the individuals' mental health and the development of relevant tools and support. As such, providing methods to implement and evaluate dimensional emotion models will further help to utilize TED systems in more complex health applications.

Advantages of clinical data for health-related TED: Using clinical data may provide insights into people's genuine emotions and may lead to a more accurate diagnosis of psychological disorders. Gathering data from medical web resources as well as from direct interactions with chatbots helps research groups to develop knowledge/data-bases to further the progress of TED in patient-facing applications.

Deep learning TED in health: Deep learning models have the potential for real-time and multi-faceted TED solutions reliant on several data modalities and sources. Recent advances in transfer learning and multitask learning processes [106,107] in deep learning, can further enhance multi-faceted TED, in combination with other closely related tasks. This makes use of the emotions of individuals to make further inferences, such as downstream decision-making toward adverse drug reactions.

Table A.6
Data sources for TED in health and medicine.

Clinical data		
Online medical healthcare platforms	Webmd[a]	[47]
	Healthtap[b]	[47]
Online therapy	Horyzon [108]	[48]
	PsychPark[c]	[49]
Clinical interviews	Distress Analysis Interview Corpus Wizard-of-Oz (DAIC-WOZ) [94]	[50]
Medical sensors embedded in smart homes/cities	Converted text from patient audio data	[51]
Non-clinical data		
Social media	Twitter	[5–8,34,36,37,40,41,52–65]
	Twitter-based datasets	
	SenWave [43]	[66]
	Covid Stance [45]	[67]
	SemEval-2018:(AIT) [44]	[66,68]
	COVID-19 Twitter Chatter [46]	[69]
	Kaggle[d]	[57]
	Emotional Tweets [109]	[57]
	Sentiment140 [110]	[41,57]
	Reddit	[7,39,70]
	Weibo	[8,35,42,68,71]
	Facebook	[53,63]
	YouTube	[53,56,60,63]
	Instagram	[53,56,60,63]
	Official press websites, news articles, internet forums and blogs	[38,56,60]
Chatbot conversations		[10,11,28,72–79]
Online health communities (OHC)	Lung and breast cancer forums	[80]
	Mental health support forums	[81]
	Facebook diabetes community	[82]
	Cancer support groups	[83]
	Autism support site in China[e]	[35]
	MEDHelp[f]	[84]
Global English news sources	Reuters, BBC, Yahoo News, South China Morning Post, National Post, Daily Mail UK, CNBC, The Guardian, CNN	[9]
Crowdsourcing and interview	Suicide notes from 2011 i2b2 NLP Challenge	[85–92]
	COVID-19 Real World Worry Dataset	[27]
	Collection of conversation-like text	[93]

[a] https://www.webmd.com/.
[b] https://www.healthtap.com/.
[c] http://www.psychpark.org.
[d] https://www.kaggle.com/smid80/coronavirus-covid19-tweets.
[e] http://www.guduzheng.ne.
[f] https://www.medhelp.org/.

Applications of TED in health other than COVID-19: A deeper look at TED applications for other diseases provides insights on how to make use of TED for medical simulations [75] or for receiving feedback from patients with respect to medical services [84], membership in health communities [83], and public awareness about diseases [85].

Potential users of TED in health: Our analysis suggests that several different groups of users currently benefit from textual emotion recognition in health and medicine: (i) individual users of social media networks and healthcare platforms who are prone to mental disorders or are in need of further support to manage the difficulties of such disorders, (ii) patients with referrals to online counseling systems, (iii) specialists and physicians who receive feedback on treatment services or use TED systems for educational purposes, and (iv) social analysts, government decision-makers, senior health planners, and supporters of health communities who analyze and plan for improved and higher quality support for patients.

4.1. Limitations and challenges

Research endeavors are often faced with limitations and challenges. Where fields of computer science and health combine, meeting industry standards in both fields presents an additional complication. The application of NLP to the domain of healthcare requires a level of understanding from the computer science research team of health delivery and associated ethical considerations and vice-versa. This limitation was observed, for example, in [76] who claim to possibly cure mental health issues, and [74] stating disclosure of feelings is the second most efficient way to curb depression in the youth. We see a potential mismatch where clinical claims are not backed with clinical research data as thoroughly as the computer science aspects of the research. One group did however acknowledge the limitation of finding issues with in-depth knowledge of the field when selecting search terms [84]. This limitation should be considered by research groups moving forwards. Field diversity of the research group or research outcomes reviewed by a member of the medical academic community is suggested, to ensure resulting technologies will be certain to do no harm.

Table B.7
Datasets used for TED method development and training.

Dataset	Description	Use case
Emotion in Text [111]	A large open-source emotion dataset published by CrowdFlower	[7]
Unified [112]	An aggregation of existing emotion classification datasets from different sources and domains	[67]
GoEmotion [113]	A manually annotated dataset of English Reddit comments with corresponding emotion labels	[67]
ISEAR [114]	The International Survey on Emotion Antecedents and Reactions. People's reports about the situations and emotions they experienced	[50,74,78]
SemEval 2007: Affective text dataset [115]	The affective text dataset of categorized sentences from different news sources on the Internet provided by the SemEval 2007 workshop	[50,80]
SemEval 2013 [116]	Tweets provided by the SemEval 2013 workshop	[34]
UCLA/UW Couple Therapy Corpus [117]	Recordings of real couples with marital issues interacting over multiple sessions. Recordings were rated by multiple annotators based on the Couples Interaction [118] and Social Support [119] rating system	[95]
IEMOCAP [120]	Interactive Emotional Dyadic Motion Capture Database: Emotion-rated recordings from five male–female pairs of actors performing both scripted and improvised dyadic interactions	[95]
Emotion dataset [121]	A dataset of English Twitter messages with six basic emotions published by Hugging Face	[78]

Table C.8
Main objectives of emotion analysis in healthcare platforms and communities in different health domains.

Deployment	Application	Health domain
Healthcare platforms	Developing chatbots and mobile applications for psychiatric counseling and therapy	Mental health [11,48,72,73,76,79], Mental health in COVID-19 [10,70,77,78], Depression [50,74]
	Developing a virtual patient to simulate patient–doctor communication	Medical simulation [75]
	Identification of emotion characteristics, intensities and transitions across three distinct mental health groups	Mental health [81]
	Extracting various negative mental health disorders	Addiction, anxiety, depression, insomnia, stress, and Obsessive Compulsive Disorder [47]
	Predictive and assistive care system in the Internet of Health ecosystem	Alzheimer [51]
	Emotion detection in suicide notes	Suicide [85–92]
Online health communities	Promoting patient-centric care by better understanding patients' perspectives	Oculoplastic surgery (eye health) [84]
	Analysis and prediction of user reactions based on their emotions	Diabetes [82]
	Investigating the impact of public holidays on emotional states in patients	Breast and lung cancer [80]
	Interpreting the emotional state of patients when joining online support groups	Prostate cancer [83]

The limitations and challenges which were identified include ethical and data security concerns where no age or gender was associated with data sources [7] leading to misrepresenting certain demographics. Models developed and trained for specific cases (i.e., COVID-19) [52], may not be able to be applied to other datasets (i.e., Ebola), and in one case, an acknowledgment that the "therapy chatbot" developed is not able to replace therapy [73]. Data limitations included limited access to training data [70,76], a small dataset with limited emotions detected [82], limited domain-specific labeled data [81], under-labeled datasets [5,95], and sparse data [92]. The implication of data limitations was reduced performance or accuracy of results, and unlabeled data inspired authors to deploy unsupervised machine learning instated of supervised techniques.

Table D.9
Methods used for TED in health and medicine.

Lexicon-based approaches	
Keyword-based methods	[5,9,10,35,36,39,42,54,58,59,75]
Corpus-based methods	[61,69]
Machine learning models	
Support Vector Machines (SVMs)	[49,55,63,73,85,86,91,92]
Growing Self-Organising Maps (GSOM)	[64,81]
Linear Regression	[27]
Logistic Regression	[34,90]
Naïve Bayes (NB)	[6,49,51,55,63]
Maximum Entropy (ME)	[87]
C4.5	[49]
K-Nearest Neighbor (KNN)	[63]
Random Forest (RF)	[63,65]
SDMs (Smoothed Dirichlet Multinomial distribution) + Likelihood-based Learning	[41]
Deep learning models	
Convolutional Neural Network (CNN)	[74]
Hierarchical CNN	[8]
Recursive Neural Network (RNN)	[28,74]
BERT	[67,78]
RoBERTa (Optimized BERT)	[7]
AraBERT, MARBERT	[66]
Description-based BERT	[68]
Long Short Term Memory (LSTM)	[11,50,53,57,67,72]
Bidirectional LSTM	[71]
BiLSTM-CRF	[70]
SAB-LSTM (extended BiLSTM)	[53]
Hierarchical Attention Network (HAN)	[74]
Hybrid models	
ML (SVMs, ME, NB) + Lexicon-based model and Conditional Random Field (CRF)	[37,38,62,89]
ML (SVMs) + Rule-based engine	[88]
PRIME Framework [122]: ML + DL	[83]
Deep Learning models combinations:	
Multi-head attention + biLSTM and CNN	[47]
CNN + lexicon-based features + LSTM	[80]
Gated Recurrent Units (GRUs) + LSTM + ARTMAP NN	[11,72,77]
Multitask learning (RNN + GRUs)	[95]
Cloud models, open-source frameworks, and services	
IBM's Watson [104]	[56,60,84]
Indico API[a]	[82]
DialogFlow[b]	[76,79]
Facebook's Fasttext	[40]
Syuzhet[c] [123] package available in R	[52]

[a] https://indico.io/docs/emotion.
[b] https://cloud.google.com/dialogflow.
[c] https://github.com/mjockers/syuzhet.

A geographical bias was identified in a number of studies. Missing geo-location data may result in over-representing particular areas and individuals [84] and a lack of location data in [58] significantly slowed processing and reduced the accuracy of their models. Lack of location data in [69] resulted in the researchers using the text: "India/Indians" to verify location and in [37] the sparseness of geographical metadata reduced the size of their text corpus. A selection bias was noted in a reluctance of individuals to use web forums to discuss health-related issues [84] where only computer users are contributing to datasets [83], and these users are aged between 18 to 40 years [42]. This bias implies that the results are unable to be generalized to the entire population.

Other limitations were related to data available only in English [39,53,58,69], with misspellings or poor grammar noted in annotated datasets significantly reducing the size of the dataset available, impacting the accuracy and generalizability of the results [85,88,90,92]. This restricts the dataset and reduces the ability to generalize the results to a larger population [6,58,67,69,71]. The inability of the models to detect sarcasm was thought to impact the classification of extracted data, as the model could not differentiate between genuine and sarcastic tones, and models being unable to process humor, idioms, or irony simply being unable to detect these more complex emotions or expressions used by humans [36,40,81,90]. Consequently, the use of these models and resulting technologies in very remote non-English speaking elderly and poorly educated populations cannot be sure to be accurate or beneficial.

4.2. Future trends

The future direction for research in health-related TED is two-fold including developments in both aspects of health and TED techniques. Research in the technical development of TED is moving in several directions aimed toward more efficient, universal models. Researchers seek to develop multi-faceted (multi-lingual, multi-modal, and multi-labeled) approaches for emotion recognition. Although work has begun in multi-lingual [55], multi-label [7,66], and multi-modal [41,50,78] approaches to TED, there is room for improvement, such as morphological analysis of text in other languages [10,66,71], extending approaches for multi-modality [67], and generation of resources (lexicons, corpora, and multi-modal datasets [129]) to achieve acceptable, enhanced performance and reliable systems.

Table E.10
List of lexicons used for TED in health.

Lexicon	Description	Reference utilizing lexicon
WordNet-Affect [96]	Includes a subset of WordNet synsets that represent affective concepts in a hierarchically organized tree structure.	[34,55]
LIWC[a] [124]	LIWC is a basic and popular tool for the collection of linguistic statistics from text. The LIWC dictionary is composed of 5,690 words and word stems. Each word or word stem defines one or more word categories.	[75]
NRC[b] Word-Emotion Association Lexicon (EmoLex)[c] [98,99]	A list of English words and their associations with eight basic emotions and two sentiments. The annotations were manually performed through crowdsourcing.	[9,36,54,54,59,61,69,69,80]
NRC Emotion Intensity Lexicon (NRC-EIL) [125]	A dictionary of 9,922 lemmas associated with scores reflecting the emotions they convey and their intensity.	[39]
DUTIR[d] [126]	A general Chinese emotion lexicon which contains 27,446 common Chinese words with seven emotion labels [127].	[35,42]
A GitHub dataset[e]	Contains synonyms of keywords corresponding to each of the six basic emotions.	[5]
Social surveillance data[f]	Contains 499 terms corresponding to nine emotion classes, including the 6-scale emotion-related terms, lexical units, emotion terms, emoticons, and "news-related" terms.	[37,38]
DepecheMood [128]	Contains (word, emotion) association scores obtained through manual annotations of news headlines.	[61]

[a]Linguistic Inquiry and Word Count.
[b]The Canadian National Research Council.
[c]http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm.
[d]http://ir.dlut.edu.cn/.
[e]https://github.com/timjurka/sentiment/blob/master/sentiment/data/emotions.csv.gz.
[f]https://bitbucket.org/readbiomed/socialsurveillance.

The second path of technical developments in TED includes high-performance computational tools and approaches [66,70], deep learning models [8,63], and semantically rich representations using an ontology for effectively detecting individuals' opinions [57].

Adding more syntactic and semantic features [37,38,70,88], utilizing transformer-based deep learning techniques [47], and using more advanced forms of text pre-processing [85,92] will improve the text classification accuracy for TED. Future advances in TED are also based on the extraction of more emotional clues for each emotion category [88] and the identification of implicit and ambiguous emotional expressions [10,89]. Modeling emotion intensities and transitions [90], context-based [81] and content-based analysis of emotion data [55], and detection of emotional triggers or reasons [39,68,76] are also current trends in the area of emotion detection.

Future research utilizing TED in health and medicine may focus on mental health by automatically identifying patients with severe emotional distress [83,85] through the analysis of emotions extracted from real-time streaming data in social media [5,40,58,60]. Developing emotional monitoring systems over specific regions of the world for disease outbreaks [37] is another under-researched avenue for further developments of TED in health.

Recognizing emotions will aid the development of medical applications, like interpreting medical signs, symptoms, and diseases, decreasing appointment and waiting times, and addressing patients' concerns regarding specific procedures [84]. Future TED-versed chatbots are likely to learn user preferences and provide feedback or suggest exercises depending on individual situations and emotions [10]. Developing chat-based pre-counseling sessions [50] for immediate clinical intervention [83] is being studied by researchers in the field. Utilizing health-related lexicons to be integrated into the parsing processes of textual data will also help to improve the effectiveness of these systems [28].

Future developments in the technical aspects of TED to improve accuracy and precision will complement the future directions of TED in health by reducing the risk of type-one and type-two errors when interacting with human emotion data. This combined area has potential for future research on the accuracy of chatbots used in mental illness identification which will improve confidence in the application of these technologies to human populations.

5. Conclusion

Textual Emotion Detection (TED) is impacting several technological advances in health and medicine. The effect of TED improving applications such as monitoring medical events on social media, developing healthcare-supporting platforms, and supporting online health communities also coincides with major developments in the NLP and machine learning domains in recent years. This article utilized the PRISMA 2020 guidelines on systematic meta-analyses and reviewed a total of 66 articles related to emotion analysis and detection in health and medicine. The information within the articles was particularly analyzed with reference to the datasets utilized for any emotion-related text classification, the specific health-related applications, the algorithm and methodology developed for TED, as well as the performance measures and outcomes reported on TED. Approximately, 44% of the articles focused on the recent COVID-19 pandemic-related emotions, and the most accurate TED systems were found to utilize current deep learning-based approaches to TED. The most common applications of TED were in public health monitoring and public reaction analysis of health-related events using massive amounts of text available on social media networks.

Although the intersection of TED and health domains has seen significant developments in recent years, this article also found some current challenges that require further investigation. Multi-lingual, multi-labeled, multi-modal TED, in-depth semantic and knowledge-based approaches to TED in combination with current deep learning advances, more advanced text pre-processing techniques, more reliable assessment strategies and studies on hybrid approaches in TED, the implementation and evaluation of dimensional emotional models, and the contribution of health-related knowledge/data-bases to further the progress of TED in patient-facing applications are among major future works. In terms of health applications, there remains room for further utilization of TED in several areas such as predicting mental health disorders, emotional monitoring of patients on specific social media platforms, and using the emotional stance of patients to further improve patient experiences as they interact with practitioners, healthcare systems, and health organizations.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A

See Table A.6.

Appendix B

See Table B.7.

Appendix C

See Table C.8.

Appendix D

See Table D.9.

Appendix E

See Table E.10.

References

[1] W. Medhat, A. Hassan, H. Korashy, Sentiment analysis algorithms and applications: A survey, Ain Shams Eng. J. 5 (4) (2014) 1093–1113, http://dx.doi.org/10.1016/j.asej.2014.04.011.
[2] P. Ekman, Facial expressions of emotion: New findings, new questions, Psychol. Sci. 3 (1) (1992) 34–38, http://dx.doi.org/10.1111/j.1467-9280.1992.tb00253.x.
[3] G. Gonzalez-Hernandez, A. Sarker, K. O'Connor, G. Savova, Capturing the patient's perspective: A review of advances in natural language processing of health-related text, IMIA Yearbook Med. Inform. 26 (1) (2017) 214–227, http://dx.doi.org/10.15265/IY-2017-029.
[4] O.G. Iroju, J.O. Olaleke, A systematic review of natural language processing in healthcare, Int. J. Inf. Technol. Comput. Sci. 8 (8) (2015) 44–50, http://dx.doi.org/10.5815/ijitcs.2015.08.07.
[5] A.S.M. Venigalla, S. Chimalakonda, D. Vagavolu, Mood of India during Covid-19-an interactive web portal based on emotion analysis of Twitter data, in: Conference Companion Publication of the 2020 on Computer Supported Cooperative Work and Social Computing, 2020, pp. 65–68.
[6] E. D'Avanzo, G. Pilato, M. Lytras, Using Twitter sentiment and emotions analysis of Google trends for decisions making, Program (2017).
[7] J. Choudrie, S. Patil, K. Kotecha, N. Matta, I. Pappas, Applying and understanding an advanced, novel deep learning approach: A covid 19, text based, emotions analysis study, Inf. Syst. Front. Jun 25 (2021) 1–35, http://dx.doi.org/10.1007/s10796-021-10152-6.
[8] R.L. Rosa, M.J. De Silva, D.H. Silva, M.S. Ayub, D. Carrillo, P.H. Nardelli, D.Z. Rodriguez, Event detection system based on user behavior changes in online social networks: Case of the COVID-19 pandemic, Ieee Access 8 (2020) 158806–158825.
[9] F. Aslam, T.M. Awan, J.H. Syed, A. Kashif, M. Parveen, Sentiments and emotions evoked by news headlines of coronavirus disease (COVID-19) outbreak, Humanit. Soc. Sci. Commun. 7 (1) (2020) 1–9.
[10] K. Denecke, S. Vaaheesan, A. Arulnathan, A mental health chatbot for regulating emotions (SERMO)-concept and usability test, IEEE Trans. Emerg. Top. Comput. 9 (3) (2020) 1170–1182.
[11] D. Lee, K.-J. Oh, H.-J. Choi, The chatbot feels you-a counseling service using emotional response generation, in: 2017 IEEE International Conference on Big Data and Smart Computing (BigComp), IEEE, 2017, pp. 437–440.
[12] S. Peng, L. Cao, Y. Zhou, Z. Ouyang, A. Yang, X. Li, W. Jia, S. Yu, A survey on deep learning for textual emotion analysis in social networks, Digit. Commun. Netw. (2021).
[13] F.A. Acheampong, C. Wenyu, H. Nunoo-Mensah, Text-based emotion detection: Advances, challenges, and opportunities, Eng. Rep. 2 (7) (2020) e12189.
[14] A.R. Murthy, K.A. Kumar, A review of different approaches for detecting emotion from text, in: IOP Conference Series: Materials Science and Engineering, Vol. 1110, (1) IOP Publishing, 2021, 012009.
[15] P. Nandwani, R. Verma, A review on sentiment analysis and emotion detection from text, Soc. Netw. Anal. Min. 11 (1) (2021) 1–19.
[16] M.V. Mäntylä, D. Graziotin, M. Kuutila, The evolution of sentiment analysis—A review of research topics, venues, and top cited papers, Comp. Sci. Rev. 27 (2018) 16–32.
[17] M. Elbattah, É. Arnaud, M. Gignon, G. Dequen, The role of text analytics in healthcare: A review of recent developments and applications, in: HEALTHINF, 2021, pp. 825–832.
[18] A. Gupta, R. Katarya, Social media based surveillance systems for healthcare using machine learning: A systematic review, J. Biomed. Inform. 108 (2020) 103500.
[19] R.A. Calvo, D.N. Milne, M.S. Hussain, H. Christensen, Natural language processing in mental health applications using non-clinical texts, Nat. Lang. Eng. 23 (5) (2017) 649–685.
[20] A. Le Glaz, Y. Haralambous, D.-H. Kim-Dufor, P. Lenca, R. Billot, T.C. Ryan, J. Marsh, J. Devylder, M. Walter, S. Berrouiguet, et al., Machine learning and natural language processing in mental health: Systematic review, J. Med. Int. Res. 23 (5) (2021) e15708.
[21] A. Zunic, P. Corcoran, I. Spasic, et al., Sentiment analysis in health and well-being: Systematic review, JMIR Med. Inform. 8 (1) (2020) e16023.
[22] M.J. Page, J.E. McKenzie, P.M. Bossuyt, I. Boutron, T.C. Hoffmann, C.D. Mulrow, L. Shamseer, J.M. Tetzlaff, E.A. Akl, S.E. Brennan, et al., The PRISMA 2020 statement: An updated guideline for reporting systematic reviews, Int. J. Surg. 88 (2021) 105906.
[23] P. Whiting, J. Savović, J.P. Higgins, D.M. Caldwell, B.C. Reeves, B. Shea, P. Davies, J. Kleijnen, R. Churchill, et al., ROBIS: a new tool to assess risk of bias in systematic reviews was developed, J. Clin. Epidemiol. 69 (2016) 225–234.
[24] M.L. McHugh, Interrater reliability: the kappa statistic, Biochem. Med. 22 (3) (2012) 276–282.
[25] A.B. Cantor, Sample-size calculations for Cohen's kappa, Psychol. Methods 1 (2) (1996) 150.
[26] M.A. Bujang, N. Baharum, Guidelines of the minimum sample size requirements for Kappa agreement test, Epidemiol. Biostat. Public Health 14 (2) (2017).
[27] B. Kleinberg, I. van der Vegt, M. Mozes, Measuring emotions in the covid-19 real world worry dataset, 2020, arXiv preprint arXiv:2004.04225.
[28] K. Denecke, R. May, Y. Deng, Towards emotion-sensitive conversational user interfaces in healthcare applications, Stud. Health Technol. Inform. 264 (2019) 1164–1168.
[29] R. Plutchik, A general psychoevolutionary theory of emotion, in: Theories of Emotion, Elsevier, 1980, pp. 3–33.
[30] A. Ortony, G.L. Clore, A. Collins, The Cognitive Structure of Emotions, Cambridge University Press, 1990.
[31] J.A. Russell, A circumplex model of affect, J. Personal. Soc. Psychol. 39 (6) (1980) 1161.
[32] X. Jin, Z. Wang, An emotion space model for recognition of emotions in spoken Chinese, in: J. Tao, T. Tan, R.W. Picard (Eds.), Affective Computing and Intelligent Interaction, Springer Berlin Heidelberg, Berlin, Heidelberg, 2005, pp. 397–402.
[33] W.G. Parrott, Emotions in Social Psychology: Essential Readings, psychology Press, 2001.
[34] S. Volkova, L.E. Charles, J. Harrison, C.D. Corley, Uncovering the relationships between military community health and affects expressed in social media, Epj Data Sci. 6 (1) (2017) 9.
[35] T.Y. Tang, R.Y. Wang, C.G. Chen, The sum of all our feelings!: Sentimental analysis on Chinese autism sites, in: International Conference on Universal Access in Human-Computer Interaction, Springer, 2017, pp. 108–116.
[36] A.D. Dubey, Decoding the twitter sentiments towards the leadership in the times of COVID-19: A case of USA and India, 2020, Available at SSRN 3588623.

[37] B. Ofoghi, M. Mann, K. Verspoor, Towards early discovery of salient health threats: A social media emotion classification technique, in: Biocomputing 2016: Proceedings of the Pacific Symposium, World Scientific, 2016, pp. 504–515.
[38] B. Ofoghi, K. Verspoor, Textual emotion classification: An interoperability study on cross-genre data sets, in: Australasian Joint Conference on Artificial Intelligence, Springer, 2017, pp. 262–273.
[39] V. Basile, F. Cauteruccio, G. Terracina, How dramatic events can affect emotionality in social posting: The impact of COVID-19 on reddit, Future Int. 13 (2) (2021) 29.
[40] A. Kulai, M. Sankhe, S. Anglekar, A. Halbe, Emotion analysis of Covid Tweets using FastText supervised classifier model, in: 2021 International Conference on Communication Information and Computing Technology, ICCICT, IEEE, 2021, pp. 1–6.
[41] F. Najar, N. Bouguila, Emotion recognition: A smoothed Dirichlet multinomial solution, Eng. Appl. Artif. Intell. 107 (2022) 104542.
[42] Y. Su, P. Wu, S. Li, J. Xue, T. Zhu, Public emotion responses during COVID-19 in China on social media: An observational study, Hum. Behav. Emerg. Technol. 3 (1) (2021) 127–136.
[43] Q. Yang, H. Alamro, S. Albaradei, A. Salhi, X. Lv, C. Ma, M. Alshehri, I. Jaber, F. Tifratene, W. Wang, et al., Senwave: Monitoring the global sentiments under the COVID-19 pandemic, 2020, arXiv preprint arXiv:2006.10842.
[44] S. Mohammad, F. Bravo-Marquez, M. Salameh, S. Kiritchenko, Semeval-2018 task 1: Affect in tweets, in: Proceedings of the 12th International Workshop on Semantic Evaluation, 2018, pp. 1–17.
[45] E.C. Mutlu, T. Oghaz, J. Jasser, E. Tutunculer, A. Rajabi, A. Tayebi, O. Ozmen, I. Garibay, A stance data set on polarized conversations on Twitter about the efficacy of hydroxychloroquine as a treatment for COVID-19, Data Brief 33 (2020) 106401.
[46] J.M. Banda, R. Tekumalla, G. Wang, J. Yu, T. Liu, Y. Ding, E. Artemova, E. Tutubalina, G. Chowell, A large-scale COVID-19 Twitter chatter dataset for open scientific research—an international collaboration, Epidemiologia 2 (3) (2021) 315–324.
[47] K. Dheeraj, T. Ramakrishnudu, Negative emotions detection on online mental-health related patients texts using the deep learning with MHA-BCNN model, Expert Syst. Appl. 182 (2021) 115265.
[48] S. D'Alfonso, O. Santesteban-Echarri, S. Rice, G. Wadley, R. Lederman, C. Miles, J. Gleeson, M. Alvarez-Jimenez, Artificial intelligence-assisted online social therapy for youth mental health, Front. Psychol. 8 (2017) 796.
[49] J.-L. Wu, P.-C. Chang, S.-L. Chang, L.-C. Yu, J.-F. Yeh, C.-S. Yang, Emotion classification by incremental association language features, in: Proceedings of World Academy of Science, Engineering and Technology, Vol. 65, Citeseer, 2010, pp. 487–491.
[50] S. Prabhu, H. Mittal, R. Varagani, S. Jha, S. Singh, Harnessing emotions for depression detection, Pattern Anal. Appl. (2021) 1–11.
[51] S. Sharma, R.K. Dudeja, G.S. Aujla, R.S. Bali, N. Kumar, Detras: deep learning-based healthcare framework for IoT-based assistance of Alzheimer patients, Neural Comput. Appl. (2020) 1–13.
[52] R. Goel, R. Sharma, Studying leaders & their concerns using online social media during the times of crisis - A COVID case study, Soc. Netw. Anal. Min. 11 (1) (2021).
[53] D.A. Kumar, A. Chinnalagu, Sentiment and emotion in social media COVID-19 conversations: SAB-LSTM approach, in: 2020 9th International Conference System Modeling and Advancement in Research Trends, SMART, IEEE, 2020, pp. 463–467.
[54] C. Crocamo, M. Viviani, L. Famiglini, F. Bartoli, G. Pasi, G. Carrá, Surveilling COVID-19 emotional contagion on Twitter, Eur. Psychiatry: J. Assoc. Eur. Psychiatrists (2021) 1–21.
[55] V.K. Jain, S. Kumar, S.L. Fernandes, Extraction of emotions from multilingual text using intelligent text processing and computational linguistics, J. Comput. Sci. 21 (2017) 316–326.
[56] P.P. Iglesias-Sánchez, G.F. Vaccaro Witt, F.E. Cabrera, C. Jambrino-Maldonado, The contagion of sentiments during the COVID-19 pandemic crisis: The case of isolation in Spain, Int. J. Environ. Res. Public Health 17 (16) (2020) 5918.
[57] A.S. Imran, S.M. Daudpota, Z. Kastrati, R. Batra, Cross-cultural polarity and emotion detection using sentiment analysis and deep learning on COVID-19 related tweets, Ieee Access 8 (2020) 181074–181090.
[58] M. Alam, M. Kaschura, H. Sack, Apollo: Twitter stream analyzer of trending hashtags: A case-study of# COVID-19, in: ISWC (Demos/Industry), 2020.
[59] J.C. Lyu, E.L. Han, G.K. Luli, Covid-19 vaccine-related discussion on twitter: Topic modeling and sentiment analysis, J. Med. Int. Res. 23 (6) (2021).
[60] C. de Las Heras-Pedrosa, P. Sánchez-Núñez, J.I. Peláez, Sentiment analysis and emotion understanding during the COVID-19 pandemic in Spain and its impact on digital ecosystems, Int. J. Environ. Res. Public Health 17 (15) (2020) 5542.
[61] S.D. Gollapalli, P. Rozenshtein, S.K. Ng, ESTeR: Combining word co-occurrences and word associations for unsupervised emotion detection, in: Findings of the Association for Computational Linguistics: EMNLP 2020, 2020, pp. 1043–1056.
[62] B. Ofoghi, S. Siddiqui, K. Verspoor, READ-Biomed-SS: Adverse drug reaction classification of microblogs using emotional and conceptual enrichment, in: PSB 2016: Social Media Mining Shared Task Workshop, [The Conference], 2016, pp. 1–5.
[63] Z. Nabeel, M. Mehmood, A. Baqir, A. Amjad, Classifying emotions in roman urdu posts using machine learning, in: 2021 Mohammad Ali Jinnah University International Conference on Computing, MAJICC, IEEE, 2021, pp. 1–7.
[64] A. Adikari, R. Nawaratne, D. De Silva, S. Ranasinghe, O. Alahakoon, D. Alahakoon, et al., Emotions of COVID-19: content analysis of self-reported information using artificial intelligence, J. Med. Int. Res. 23 (4) (2021) e27341.
[65] D. Wehbe, A. Alhammadi, H. Almaskari, K. Alsereidi, H. Ismail, UAE e-learning sentiment analysis framework, in: The 7th Annual International Conference on Arab Women in Computing in Conjunction with the 2nd Forum of Women in Research, 2021, pp. 1–4.
[66] N. Alturayeif., H. Luqman, Fine-grained sentiment analysis of arabic covid-19 tweets using BERT-based transformers and dynamically weighted loss function, Appl. Sci. (Switzerland) 11 (22) (2021).
[67] R. Kumari, N. Ashok, T. Ghosal, A. Ekbal, Misinformation detection using multitask learning with mutual learning for novelty detection and emotion recognition, Inf. Process. Manage. 58 (5) (2021) 102631.
[68] X. Li, M. Zhou, J. Wu, A. Yuan, F. Wu, J. Li, Analyzing COVID-19 on online social media: Trends, sentiments and emotions, 2020, arXiv preprint arXiv:2005.14464.
[69] S. Das, A. Dutta, Characterizing public emotions and sentiments in COVID-19 environment: A case study of India, J. Hum. Behav. Soc. Environ. 31 (1–4) (2021) 154–167.
[70] I. Salhi, K. El Guemmat, M. Qbadou, K. Mansouri, Towards developing a pocket therapist: An intelligent adaptive psychological support chatbot against mental health disorders in a pandemic situation, Indonesian J. Electr. Eng. Comput. Sci. 23 (2) (2021) 1200–1211.
[71] X. Hu, J. Shu, Z. Jin, Depression tendency detection model for weibo users based on Bi-LSTM, in: 2021 IEEE International Conference on Artificial Intelligence and Computer Applications, ICAICA, 2021, pp. 785–790, http://dx.doi.org/10.1109/ICAICA52286.2021.9497931.
[72] K.-J. Oh, D. Lee, B. Ko, J. Hyeon, H.-J. Choi, Empathy bot: Conversational service for psychiatric counseling with chat assistant, in: MEDINFO 2017: Precision Healthcare Through Informatics, IOS Press, 2017, p. 1235.
[73] H. Shah, P. Anilkumar, S. Sakpal, N. Pathak, EDRA – An emotional health detection and recognition assistant, in: 2021 Asian Conference on Innovation in Technology, ASIANCON, 2021, pp. 1–6, http://dx.doi.org/10.1109/ASIANCON51346.2021.9544973.
[74] F. Patel, R. Thakore, I. Nandwani, S.K. Bharti, Combating depression in students using an intelligent chatbot: A cognitive behavioral therapy, in: 2019 IEEE 16th India Council International Conference, INDICON, IEEE, 2019, pp. 1–4.
[75] T.B. Talbot, M. Hackett, W. Pike, Perceptive patient: Important factors for practical emotion sensing in conversational human-computer, in: International Conference on Applied Human Factors and Ergonomics, Springer, 2021, pp. 239–246.
[76] V. Dhanasekar, Y. Preethi, S. Vishali, I.R. Praveen Joe, M.B. Poolan, A chatbot to promote students mental health through emotion recognition, in: Third International Conference on Inventive Research in Computing Applications, ICIRCA, 2021, pp. 1412–1416, http://dx.doi.org/10.1109/ICIRCA51532.2021.9544838.
[77] K.-J. Oh, D. Lee, B. Ko, H.-J. Choi, A chatbot for psychiatric counseling in mental healthcare service based on emotional dialogue analysis and sentence generation, in: 2017 18th IEEE International Conference on Mobile Data Management, MDM, IEEE, 2017, pp. 371–375.
[78] A. Vishwakarma, S. Sawant, P. Sawant, R. Shankarmani, An emotionally aware friend: Moving towards artificial general intelligence, in: Third International Conference on Inventive Research in Computing Applications, ICIRCA, 2021, pp. 1094–1100, http://dx.doi.org/10.1109/ICIRCA51532.2021.9544616.
[79] A. Podrazhansky, H. Zhang, M. Han, S. He, A chatbot-based mobile application to predict and early-prevent human mental illness, in: Proceedings of the 2020 ACM Southeast Conference, 2020, pp. 311–312.
[80] H. Khanpour, C. Caragea, Fine-grained emotion detection in health-related online posts, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018, pp. 1160–1166.
[81] A. Adikari, G. Gamage, D. de Silva, N. Mills, S.-M.J. Wong, D. Alahakoon, A self structuring artificial intelligence framework for deep emotions modeling and analysis on the social web, Future Gener. Comput. Syst. 116 (2021) 302–315, http://dx.doi.org/10.1016/j.future.2020.10.028.
[82] V. Balakrishnan, V. Govindan, N.I. Arshad, L. Shuib, E. Cachia, Facebook user reactions and emotion: An analysis of their relationships among the online diabetes community, Malaysian J. Comput. Sci. (2019) 87–97.
[83] A. Adikari, D. De Silva, W.K. Ranasinghe, T. Bandaragoda, O. Alahakoon, R. Persad, N. Lawrentschuk, D. Alahakoon, D. Bolton, Can online support groups address psychological morbidity of cancer patients? An artificial intelligence based investigation of prostate cancer trajectories, PLoS One 15 (3) (2020) e0229361.
[84] A.X.-L. Nguyen, X.-V. Trinh, S.Y. Wang, A.Y. Wu, Determination of patient sentiment and emotion in ophthalmology: Infoveillance tutorial on web-based health forum discussions, J. Med. Int. Res. 23 (5) (2021).
[85] B. Desmet, V. Hoste, Emotion detection in suicide notes, Expert Syst. Appl. 40 (16) (2013) 6351–6358.

[86] C. Cherry, S.M. Mohammad, B. De Bruijn, Binary classifiers and latent sequence models for emotion detection in suicide notes, Biomed. Inf. Insights 5 (2012) BII–S8933.
[87] R. Wicentowski, M.R. Sydes, Emotion detection in suicide notes using maximum entropy classification, Biomed. Inf. Insights 5 (2012) BII–S8972.
[88] A. Nikfarjam, E. Emadzadeh, G. Gonzalez, A hybrid system for emotion extraction from suicide notes, Biomed. Inf. Insights 5 (2012) BII–S8981.
[89] H. Yang, A. Willis, A. De Roeck, B. Nuseibeh, A hybrid model for automatic emotion recognition in suicide notes, Biomed. Inf. Insights 5 (2012) BII–S8948.
[90] E. Yeh, W. Jarrold, J. Jordan, Leveraging psycholinguistic resources and emotional sequence models for suicide note emotion annotation, Biomed. Inf. Insights 5 (2012) BII–S8979.
[91] J. Read, E. Velldal, L. Øvrelid, Labeling emotions in suicide notes: Cost-sensitive learning with heterogeneous features, Biomed. Inf. Insights 5 (2012) BII–S8930.
[92] B. Desmet, V. Hoste, Combining lexico-semantic features for emotion classification in suicide notes, Biomed. Inf. Insights 5 (2012) BII–S8960.
[93] S. Avveduto, N. Marchesini, G. Rubbia, How are you? Impressions on Covid-19 lockdown from women scientists in Italy, in: 4th International Conference on Gender Research, ICGR 2021, 2021, pp. 35–44.
[94] J. Gratch, R. Artstein, G. Lucas, G. Stratou, S. Scherer, A. Nazarian, R. Wood, J. Boberg, D. DeVault, S. Marsella, et al., The distress analysis interview corpus of human and computer interviews, in: Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), 2014, pp. 3123–3128.
[95] S.-Y. Tseng, B. Baucom, P. Georgiou, Unsupervised online multitask learning of behavioral sentence embeddings, PeerJ Comput. Sci. 5 (2019) e200.
[96] C. Strapparava, A. Valitutti, WordNet affect: An affective extension of WordNet, in: Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04), European Language Resources Association (ELRA), Lisbon, Portugal, 2004.
[97] G.A. Miller, WordNet: A lexical database for English, Commun. ACM 38 (11) (1995) 39–41.
[98] S.M. Mohammad, P.D. Turney, Crowdsourcing a word-emotion association lexicon, Comput. Intell. 29 (3) (2013) 436–465.
[99] S. Mohammad, P. Turney, Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon, in: Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches To Analysis and Generation of Emotion in Text, Association for Computational Linguistics, Los Angeles, CA, 2010, pp. 26–34, URL https://aclanthology.org/W10-0204.
[100] T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space, 2013, arXiv preprint arXiv:1301.3781.
[101] J. Pennington, R. Socher, C.D. Manning, Glove: Global vectors for word representation, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP, 2014, pp. 1532–1543.
[102] A. Joulin, E. Grave, P. Bojanowski, T. Mikolov, Bag of tricks for efficient text classification, 2016, arXiv preprint arXiv:1607.01759.
[103] M.E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, L. Zettlemoyer, Deep contextualized word representations, in: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), Association for Computational Linguistics, New Orleans, Louisiana, 2018, pp. 2227–2237, http://dx.doi.org/10.18653/v1/N18-1202.
[104] R.E. Hoyt, D.H. Snider, C.J. Thompson, S. Mantravadi, IBM watson analytics: Automating visualization, descriptive, and predictive statistics, JMIR Public Health Surveillance 2 (2) (2016) e5810.
[105] L. Zhang, S. Wang, B. Liu, Deep learning for sentiment analysis: A survey, Wiley Interdiscip. Rev.: Data Min. Knowl. Discov. 8 (4) (2018) e1253.
[106] R. Collobert, J. Weston, A unified architecture for natural language processing: Deep neural networks with multitask learning, in: Proceedings of the 25th International Conference on Machine Learning, ICML '08, Association for Computing Machinery, New York, NY, USA, 2008, pp. 160—167, http://dx.doi.org/10.1145/1390156.1390177.
[107] R. Caruana, Multitask learning, Mach. Learn. 28 (1) (1997) 41–75.
[108] M. Alvarez-Jimenez, S. Bendall, R. Lederman, G. Wadley, G. Chinnery, S. Vargas, M. Larkin, E. Killackey, P. McGorry, J. Gleeson, On the HORYZON: Moderated online social therapy for long-term recovery in first episode psychosis, Schizophr. Res. 143 (1) (2013) 143–149.
[109] S.M. Mohammad, F. Bravo-Marquez, WASSA-2017 shared task on emotion intensity, 2017, arXiv preprint arXiv:1708.03700.
[110] A. Go, R. Bhayani, L. Huang, Twitter sentiment classification using distant supervision, CS224N Project Report, Stanford, 1, (12) 2009, p. 2009.
[111] S. Mohammad, S. Kiritchenko, Understanding emotions: A dataset of tweets to study interactions between affect categories, in: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), European Language Resources Association (ELRA), Miyazaki, Japan, 2018, pp. 198–209, URL https://aclanthology.org/L18-1030.
[112] L.A.M. Oberländer, R. Klinger, An analysis of annotated corpora for emotion classification in text, in: Proceedings of the 27th International Conference on Computational Linguistics, 2018, pp. 2104–2119.
[113] D. Demszky, D. Movshovitz-Attias, J. Ko, A. Cowen, G. Nemade, S. Ravi, Goemotions: A dataset of fine-grained emotions, 2020, arXiv preprint arXiv:2005.00547.
[114] K.R. Scherer, H.G. Wallbott, Evidence for universality and cultural variation of differential emotion response patterning, J. Personal. Soc. Psychol. 66 (2) (1994) 310.
[115] C. Strapparava, R. Mihalcea, Semeval-2007 task 14: Affective text, in: Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007), 2007, pp. 70–74.
[116] P. Nakov, et al., SemEval-2013 task 2: Sentiment analysis in Twitter, 2nd joint onf. lexical and computational semantics (* SEM), 7th int, in: Workshop SemEval, Atlanta, Une, 2013, pp. 14–15.
[117] A. Christensen, D.C. Atkins, S. Berns, J. Wheeler, D.H. Baucom, L.E. Simpson, Traditional versus integrative behavioral couple therapy for significantly and chronically distressed married couples, J. Consult. Clin. Psychol. 72 (2) (2004) 176.
[118] C. Heavey, D. Gill, A. Christensen, Couples Interaction Rating System 2 (CIRS2), Vol. 7, University of California, Los Angeles, 2002.
[119] J. Jones, A. Christensen, Couples Interaction Study: Social Support Interaction Rating System, Vol. 7, University of California, Los Angeles, 1998.
[120] C. Busso, M. Bulut, C.-C. Lee, A. Kazemzadeh, E. Mower, S. Kim, J.N. Chang, S. Lee, S.S. Narayanan, IEMOCAP: Interactive emotional dyadic motion capture database, Lang. Resour. Eval. 42 (4) (2008) 335–359.
[121] E. Saravia, H.-C.T. Liu, Y.-H. Huang, J. Wu, Y.-S. Chen, Carer: Contextualized affect representations for emotion recognition, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018, pp. 3687–3697.
[122] W. Ranasinghe, T. Bandaragoda, D. De Silva, D. Alahakoon, A novel framework for automated, intelligent extraction and analysis of online support group discussions for cancer related outcomes, BJU Int. 120 (2017) 59–61.
[123] M.L. Jockers, Syuzhet: Extract sentiment and plot arcs from text, 2015, Retrieved October 21.
[124] Y.R. Tausczik, J.W. Pennebaker, The psychological meaning of words: LIWC and computerized text analysis methods, J. Lang. Soc. Psychol. 29 (1) (2010) 24–54.
[125] S.M. Mohammad, Word affect intensities, 2017, arXiv preprint arXiv:1704.08798.
[126] L. Xu, H. Lin, Y. Pan, H. Ren, J. Chen, Constructing the affective lexicon ontology, J. China Soc. Sci. Tech. Inf. 27 (2) (2008) 180–185.
[127] S. Wen, X. Wan, Emotion classification in microblog texts using class sequential rules, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 28, (1) 2014, pp. 187–193.
[128] J. Staiano, M. Guerini, Depechemood: A lexicon for emotion analysis from crowd-annotated news, 2014, arXiv preprint arXiv:1405.1605.
[129] Y. Jiang, W. Li, M.S. Hossain, M. Chen, A. Alelaiwi, M. Al-Hammadi, A snapshot research and implementation of multimodal information fusion for data-driven emotion recognition, Inf. Fusion 53 (2020) 209–221.