

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2021.DOI

# Case Studies on the Use of Sentiment Analysis to Assess the Effectiveness and Safety of Health Technologies: A Scoping Review

JULIE POLISENA<sup>1</sup>, MARTINA ANDELLINI<sup>1,2</sup>, PIERGIORGIO SALERNO<sup>3</sup>, SIMONE BORSCI<sup>1,4</sup>, LEANDRO PECCHIA<sup>1,2</sup>, and ERNESTO IADANZA<sup>1,3</sup>

<sup>1</sup>International Federation for Medical and Biological Engineering - Health Technology Assessment Division

<sup>2</sup>School of Engineering, University of Warwick, (UK)

<sup>3</sup>Department of Information Engineering - University of Florence, via di Santa Marta, 3 - 50139 Florence (I)

<sup>4</sup>Department of Learning, Data analysis, and Technology, University of Twente, Enschede, (NL)

Corresponding author: Ernesto Iadanza (e-mail: [ernesto.iadanza@unifi.it](mailto:ernesto.iadanza@unifi.it)).

This paragraph of the first footnote will contain support information, including sponsor and financial support acknowledgment. For example, "This work was supported in part by the U.S. Department of Commerce under Grant BS123456."

## ABSTRACT

A health technology assessment (HTA) is commonly defined as a multidisciplinary approach used to evaluate medical, social, economic, and ethical issues related to the use of a health technology in a systematic, transparent, unbiased, robust manner. To help inform HTA recommendations, the surveillance of social media platforms can provide important insights to the clinical community and to decision makers on the effectiveness and safety of the use of health technologies on a patient. A scoping review of the published literature was performed to gain some insight on the accuracy and automation of sentiment analysis (SA) used to assess public opinion on the use of health technologies. A literature search of major databases was conducted. The main search concepts were SA, social media, and patient perspective. Among the 1,776 unique citations identified, 12 studies that described the use of SA methods to evaluate public opinion on or experiences with the use of health technologies as posted on social media platforms were included. The SA methods used were either lexicon- or machine learning-based. Two studies focused on medical devices, three examined HPV vaccination, and the remaining studies targeted drug therapies. Due to the limitations and inherent differences among SA tools, the outcomes of these applications should be considered exploratory. The results of our study can initiate discussions on how the automation of algorithms to interpret public opinion of health technologies should be further developed to optimize the use of data available on social media.

**INDEX TERMS** Health Technology Assessment, HTA, Sentiment Analysis, Health Technologies, Medical Devices, Biomedical Engineering, Clinical Engineering

## I. BACKGROUND

A health technology assessment (HTA) is commonly defined as a multidisciplinary approach used to evaluate medical, social, economic, and ethical issues related to the use of a health technology in a systematic, transparent, unbiased, robust manner [1]. In HTA, efficacy refers to the benefit of using a health technology for a specific condition in a controlled setting that typically involve patients that meet a set criteria. Effectiveness, on the other hand, refers to the

benefit of using a technology for a particular condition under routine care [2].

Although randomized controlled trials (RCTs) are deemed to be the gold standard to measure the efficacy of a health technology (e.g., drug therapies), real-world data can be used to increase the efficiency of clinical trials. The use of Real World Data (RWD) can then be used to generate Real World Evidence (RWE) to assess the effectiveness of a health technology in a real-world setting.

The Food and Drugs Act defines RWD as the data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources (e.g., data collected from data registries, electronic health records, etc.), and RWE is the clinical evidence regarding the usage and potential benefits or risks of a medical product derived from analysis of RWD (e.g., information derived from multiple RWD sources, [3]).

Sources of RWD can include patient registries, collections of electronics health records (EHRs), administrative and medical claims databases, and social media platforms (e.g., Twitter®, Facebook®, blogs, etc.). Data mining, including text mining, of patient, caregiver, or health care provider opinions or experiences available on social media can provide some evidence on the effectiveness or safety of a health technology in a real-world setting.

Sentiment analysis (SA) uses natural language processing (NLP), computational linguistics, information retrieval and data mining techniques to determine the emotional tone and positive and negative opinions in a body of free-text. Organizations use SA to determine and categorize opinions about a product, service or idea. Although computerized software tools have facilitated the processing of high volumes of free-text comments into quantitative sentiment scores in a shorter timeframe, the accuracy and automation of SA remains a challenge due to the subjective and complex language use [4].

The two main approaches for SA are lexicon-based and machine-learning (ML)-based methods. The lexicon-based approach uses a dictionary or bag of words that are either positive or negative together with their corresponding polarity measure, whereas the ML-based approach builds a classification tree that encodes information to measure the variation in public opinion about the topic at hand (e.g., patient experience with the treatments received, [5]). The ML-based approach includes different methods of opinion mining e.g., Naïve Bayes, Maximum Entropy and Support Vector Machines [6], [7], [8]. Related to the ML-based approaches, deep learning techniques are emerging as powerful advancements to serve the scope of SA. These methods use neural networks (e.g. Recurrent Neural Networks with Long Short-Term Memory and Convolutional Neural Networks, etc.) to produce more accurate results than classic ML techniques, but they require more effort for algorithms training [9]. Lexicon-based and ML-Based methods are not mutually exclusive and could be coupled together to extrapolate different information from the same data-set. The ability to gather information from opinion mining throughout social media platforms can provide important insights to the clinical community and to decision-makers on the effectiveness and safety of the use of health technologies on a patient. To date, we were unable to identify any published literature on the methods or techniques used for SA on social media platforms that measure the patient, caregiver, or health care provider experiences with a health technology as part of the treatment care pathway.

To better understand the level of accuracy and automation of SA used to measure the opinions and experiences of patients, caregivers, and health care providers with the use of health technologies, a scoping review of the published literature has been conducted. Our study objectives are two-fold: i) to identify the methods/techniques used to measure free-text available on social on the use of health technologies as part of patient care and ii) to review and compare the methods and techniques used for SA as described in the selected studies. The findings can help to identify relevant methods/techniques used for sentiment analyses to inform the development of HTA recommendations.

## II. METHODS

This protocol was developed *a priori* and was followed throughout the conduct of the scoping review.

### A. LITERATURE SEARCH METHODS

The literature search was performed by an information specialist. Information was identified by searching the following bibliographic databases through the Ovid interface: MEDLINE (1946–20 August, 2019) with In-Process records and daily updates and Embase (1974 – 20 August, 2019). The search strategy was comprised of both controlled vocabulary, such as the National Library of Medicine’s MeSH (Medical Subject Headings), and keywords. The main search concepts were SA, social media and patient perspective. Citation retrieval was limited to English language documents added to the databases since January 1, 2014 until August 20, 2019. Conference abstracts were excluded from the search results. The original search strategy was updated using the same databases to capture citations from the period August 2019 up to November 20, 2020 (Supplementary File 1)

### B. SELECTION CRITERIA

The selection criteria includes articles that presented a case scenario on the application of a SA method/technique to assess the safety and effectiveness of health technologies. More specifically, studies that described the use of SA methods/techniques to evaluate the opinions of patients, caregivers, or health care providers on or their experiences with the use of health technologies as posted on social media platforms were selected for inclusion. In our review, health technologies encompass drug therapies, medical devices, medical and surgical procedures, diagnostic tests, and vaccines. Publications that discussed SA methods or techniques only but did not present a health technology-related case scenario of their application were not considered for inclusion.

### C. SCREENING AND SELECTING STUDIES FOR INCLUSION

In alignment with the scoping review protocol by [10], two reviewers (P.S. and M.A.) independently screened titles and abstracts of all citations retrieved from the literature

search according to the selection criteria. The full texts of all citations deemed to be potentially eligible by either reviewer were retrieved. The reviewers then independently reviewed the full texts, using the same selection criteria and compared their list of included and excluded studies. Any disagreements were resolved through discussion until consensus was reached, involving a third reviewer when necessary [10], [11]. Documents deemed to be eligible by both reviewers, with or without third-party adjudication, were included [11]. Reviewers used Microsoft Excel to facilitate title and abstract screening, as well as full-text study selection. The study selection process is presented in a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) extension for scoping reviews flow chart [12].

#### D. DATA COLLECTION AND ABSTRACTION

A standardized data abstraction form was used to extract data from the selected studies. Information extracted included the study characteristics (e.g., first author name, year and country of publication, objective(s), name of SA methods applied, and data sources and extraction method(s) used, health technology(ies), and stakeholder perspective(s)). Additional information extracted includes a description of the methods or techniques used to conduct a SA of the opinions on or experiences with health technologies, and the strengths and limitations of the methods/techniques as described by the authors. Finally, the population, intervention, comparator(s), outcomes, study setting, methods or technique used to describe its application, the findings, and overall conclusions were extracted from the case study. Data abstraction was performed by two reviewers (M.A. and P.S.). The data abstraction form was piloted on a random sample of two to three included articles, and modified as required. To ensure data accuracy, a third reviewer (J.P.) verified all changes made by the two reviewers.

#### E. METHODOLOGICAL QUALITY ASSESSMENT

A formal quality assessment or critical appraisal of the included articles was not conducted since our primary objective was to identify and describe the methods or techniques employed to conduct sentiment analyses on the patient opinions on or experiences with their treatments and not to test a hypothesis. This approach is aligned with the guidance on scoping review conduct [10], [11] [13], [14].

#### F. DATA ANALYSIS

One reviewer (J.P.) conducted a descriptive analysis of the methods/techniques used to carry out a SA, and a second reviewer verified the results (M.A.). The data extracted were reviewed, categorized, and organized to synthesize common methodologies. The results were then compared and interpreted to identify underlying themes and patterns from the SA methods or techniques described in the included studies.

### III. RESULTS

#### A. RESEARCH QUANTITY AVAILABLE

The search strategy yielded 1,776 articles. After duplicates were removed, a total of 1,758 citations were reviewed. Following the screening of titles and abstracts, 1,484 citations were excluded and 77 potentially relevant reports from the electronic search were retrieved for full-text review. Of these potentially relevant articles, 65 publications were excluded for various reasons (Figure 1), while 12 publications met the inclusion criteria and were included in this scoping review. Although six publications in one review were already included in our study, the remaining studies did not meet our selection criteria in terms of publication date or were focused mainly on sentiment analysis methods.

#### B. SUMMARY OF STUDY CHARACTERISTICS

##### 1) Year of Publication and Country of Origin

The studies were published between 2015 [15] and 2020 [16], [17] as shown in Figure 2a. In line with the location of corresponding authors, half of the studies (n=6/12) were written in the US [18], [19], [20], [21], [22], [23], one each in the UK [16], in Spain [24], Netherlands [25], Sweden [15], China [17], and Italy [26] (Figure 2b).

##### 2) Study Objectives

Four studies assessed and compared patient perceptions on treatments that they received for their indication [19], [21], [26], [15], four studies aimed to examine opinions on HPV vaccination available on social media [18], [16], [23],[17], and two studies were interested to learn about the patient experiences of orthodontic treatments via testimonials [25], [22]. Jiménez-Zafra sought to examine both patient and physician opinions on drug therapies expressed in user forums, and De Silva investigated the impact on influence from social media on the patient's treatment selection, experience, and recovery for prostate cancer [24], [20] (see Table I).

##### 3) Name of Sentiment Analysis Methods Applied

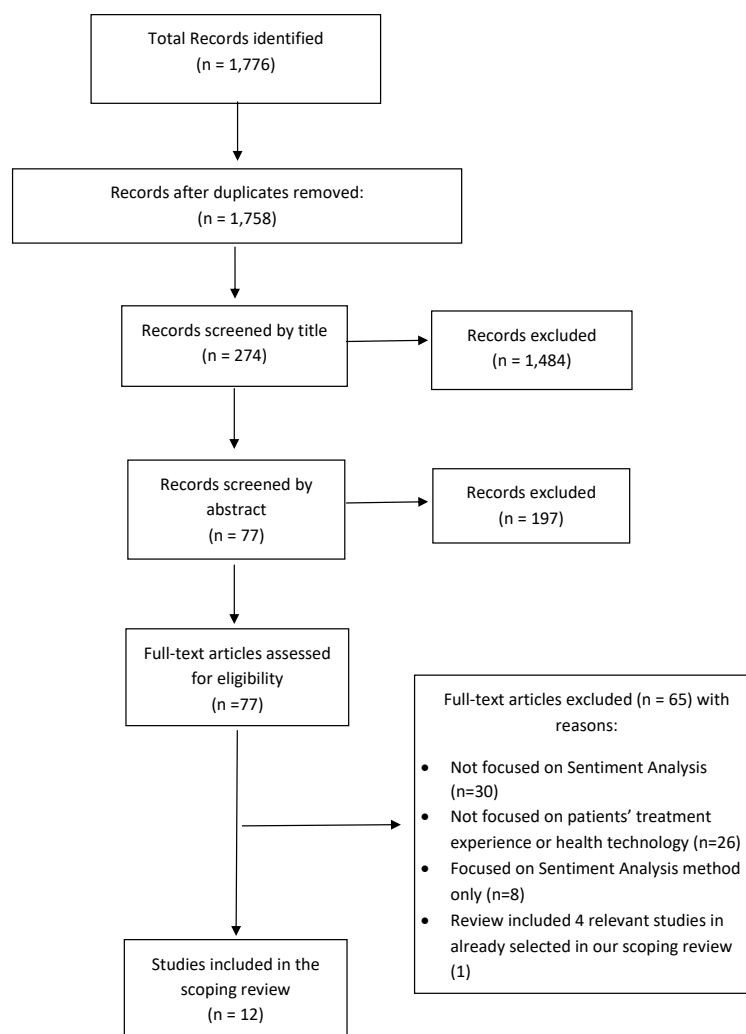
As shown in Figure 2c, four studies applied a lexicon-based approach [18], [19], [25], [21], while seven used a ML-based approach [16], [20], [22], [26], [23], [15], [17]. Jiménez-Zafra used both approaches to measure public and physician opinions on drug therapies posted in user forums [24].

#### C. DATA SOURCES

The three data sources identified in the selected studies were Twitter [18], [16], [19], [22], [23], blogs and forums [24], [20], [21], [15], YouTube [25], Facebook [26], and posts on a website (Weibo.com) [17].

##### 1) Data Extraction Method(s) Used

Du, Luo, and Wang used Twitter Application Programming Interface (API), and Zhang used both Tweepy and Twitter API to extract data from social media [23], [18],[19]. Another method used was Rapidminer [15]. The remaining



**FIGURE 1:** Selection of included studies

studies did not specify the data extraction method used (Figure 2d).

## 2) Health Technology(ies)

As shown in Figure 2e, four studies focused on cancer treatments [15], [22], [19], [20], four centred on human papillomavirus (HPV) vaccines [18], [16], [23], [17] two were on orthodontics [25], [22], and infliximab for Crohn's disease [26]. Jiménez-Zafra examined how individuals expressed their opinion on drug treatments in medical forums [24].

## 3) Stakeholder Perspective

Half of the studies analyzed patient or consumers opinions as part of the scope (n=6/12). In addition to patients, two studies also included physician or health care worker opinions in their study [24], [19]. In terms of HPV vaccines, Du and Luo focused on public opinion, and Burdens on opinions of gay, bisexual and other men who have sex with men. [23], [18], [16]. Finally, Wang focused on government opinion [17] (Figure 2f).

**TABLE I:** Study characteristics

Reference	Study objective(s)	Sentiment analysis methods	Data sources	Data extraction method(s)	Health technology(ies)	Stakeholder spective(s)
[16]	To quantify HPV vaccine messaging addressing GBpMSM and describes messaging by vaccine sentiment and characteristics.	ML-based approach	social media (i.e., Twitter)	Twitter API	HPV vaccine	GBpMSM
[17]	To explore Chinese women's perceptions of both domestic and imported HPV vaccines, which differ in price and valency.	ML-based approach for text mining	social media (Sina Weibo.com, the Chinese Twitter)	NA	HPV Vaccine	Government
[18]	To describe framework to investigate the opinions on HPV vaccination reflected on Twitter.	Lexicon-based approach	social media (i.e., Twitter)	Twitter API	HPV vaccine	General public
[24]	To examine how people express their opinion in medical forums analyzing the language used in order to determine the best way to tackle SA in this domain.	ML-based approach and lexicon-based approach	Blogs and forums (i.e., DOS [Spanish corpus of drug reviews] and COPOS [Spanish corpus of patients' opinions about physicians])	NA	Treatments based on opinions about drugs and physicians	Patients and physicians
[19]	To assess and compare perceptions about chemotherapy of patients and health care providers through analysis of chemo-related tweets.	Lexicon-based approach	Social media (i.e., Twitter)	Tweeepy and Twitter API	Chemotherapy and treatment	Patients health providers
[25]	To investigate the popularity, content of Invisalign patient testimonials on YouTube, as well as the sentiment of the related comments.	Lexicon-based approach	Social media (i.e., YouTube)	NA	Invisalign device treatment	Patients
[20]	To investigate the impact of online social influences on the intimate decision scenario of selecting a treatment type, recovery after treatment, side effects and emotions expressed over time, using prostate cancer as a model.	Semi-supervised ML combined with the Lexicon approach	Blogs and forums (i.e., OSGs)	NA	Prostate cancer treatment	Patients
[21]	To understand users' emotions and opinions on a breast cancer online support group regarding Tamoxifen	Lexicon-based approach	Blogs and forum (i.e., OSGs)	NA	Tamoxifen	Patients

Table I (continued): Study characteristics

Reference	Study objective(s)	Sentiment analysis methods	Data sources	Data extraction method(s)	Health technology(ies)	Stakeholder perspective(s)
[22]	To examine orthodontic patient experience with braces compared with that of Invisalign by means of a large-scale Twitter analysis.	ML-based approach	Social media (i.e., Twitter)	NA	Orthodontic technologies, including the Invisalign device	Patients
[26]	To comparatively investigate on infliximab's effects based on the scores obtained from: (i) a sentiment analysis returned by a SA tool (ii) an evaluation of the posts provided by human experts (i.e., gastroenterologists and non-medical experts)	ML-based approach	Social media (i.e., Facebook)	NA	Infliximab for Crohns disease	Patients
[23]	To analyze the public opinion on HPV vaccines in order to understand the reasons behind the low vaccine coverage and come up with corresponding strategies to improve vaccine uptake. The aim is to propose a machine learning system that is able to extract comprehensive public sentiment on HPV vaccines on Twitter with satisfying performance.	ML-based approach	Social media (i.e., Twitter)	Twitter API	HPV vaccine	General public
[15]	To ascertain user opinion of cancer treatment using a two-step analysis framework that focuses on positive and negative sentiment, as well as the side effects of treatment, in users' forum posts	ML-based approach	Blogs and forums (i.e., OSGs)	Rapidminer (i.e., a data collection and processing tree)	Erlotinib (Tarceva®)	Patients

a: API=Application programming interface; GBpMSM =Gay, bisexual and other men who have sex with men; HPV=Human papillomavirus; ML=Machine learning; OSGs=Online support groups; XML=Extensible Markup Language.

## D. SUMMARY OF METHODS OR TECHNIQUES USED FOR SA

### 1) Overview of Methods/Techniques

#### a: Machine Learning-Based Methods:

One study used a linear classifier to categorize the binary characteristic variables and a decision tree to classify the categorical variables [16], and De Silva developed a new ML technique based on the Emotion Wheel to capture a multi-dimensional representation of emotions expressed by patients in user forums on prostate cancer treatments [20] (Supplementary File 2). Other ML-based techniques employed to breakdown text into blocks to classify them into categories were a hierarchical Naïve Bayes SA classifier [22], OpinionFinder for natural language processing applications [26], a supervised learning model with associated learning algorithms that analyze data used for classification and regression analysis (i.e., support vector machine, [23]); the self-organizing maps (SOMs) toolbox, was used to transform the forum posts into wordlists, Wang et al. applied the Continuous Bag of Words structures and the Long Short Term Memory model to classify emotions [17], and Natural Language Toolkit (NLTK) was employed for the analysis followed by a classification of words [15].

#### b: Lexicon-Based Methods:

Luo applied the Google Cloud SA that, for a specific text, produces sentiment scores and magnitude values. The scores range from  $-1$  to  $1$ , where  $-1$  is extremely negative,  $0$  is neutral, and  $1$  is extremely positive. Furthermore, each score is linked to a magnitude value that indicates the strength of the sentiment [18]. In other studies, software based on the lexicon-based method was used to determine the overall sentiments (i.e., positive, neutral, or negative) of patients or the general public towards the health technology(ies) in question based on their opinions available on social media. They are as follows: Textblob [19], Sentistrength version 2.2 [25] and CasualConc and Linguistic Inquiry Word Count (LWIC) [21].

### 2) Combination of Lexicon- and Machine Learning-Based Methods:

Jiménez-Zafra applied both the ML supervised and lexicon-based SA approaches to evaluate patient and physician opinion extracted from social media on drug therapies [24]. The ML-based approach focused on sentiment vector machine, and the lexicon-based methods used a sentiment lexicon to find positive and negative words in the review and assigned a polarity measure to the review [24].

### 3) Strengths of Methods/Techniques as Described by the Authors

#### a: Machine Learning-Based Method:

The ML-based method used in the study by De Silva enabled the use of linear algebra to capture different semantic relationships within word-vectors in the word-embedding.

This technique can facilitate the investigation, analyses, and identification of actionable insights from patient-reported information on prostate cancer and other indications to support patient-focused healthcare delivery [20]. As noted in one study on orthodontic devices, the “context-aware” feature in the Naïve Bayes classifying technique used to extract words reduced the risk of low predictive values. As a result, there was a strong agreement in sorting by the program compared with the manual human sorting [22]. Du observed that the hierarchical classification method outperformed the plain method significantly on overall performance and for each category. The study results demonstrated the necessity of multi-classification tasks and power optimization on a corpus of tweets corpus relevant to HPV vaccinations [23]. The use of SOMs used to map large dimensional data onto a lower dimensional space accompanied by NLTK for the analysis and the classification of words enabled the identification of potential side effects consistently discussed by groups of users. This approach can serve as risk-management tool that consumers can use to express their opinion directly to the manufacturer in real-time and, subsequently, allow the manufacturers to rapidly address any problems reported [15].

#### b: Lexicon-Based Method:

The developed NLP framework in Luo’s study allows the analysis of the tweet sentiments, the extraction of key phrases, and assessment of the phrases derived from the negative tweets on HPV vaccination. This method can facilitate the investigation of HPV vaccination uptake across jurisdictions [18]. SentiStrength software enabled the extraction and processing of both positive and negative sentiments contained in textual statements. Moreover, Livas commented that SentiStrength outperformed other lexical classifiers [25].

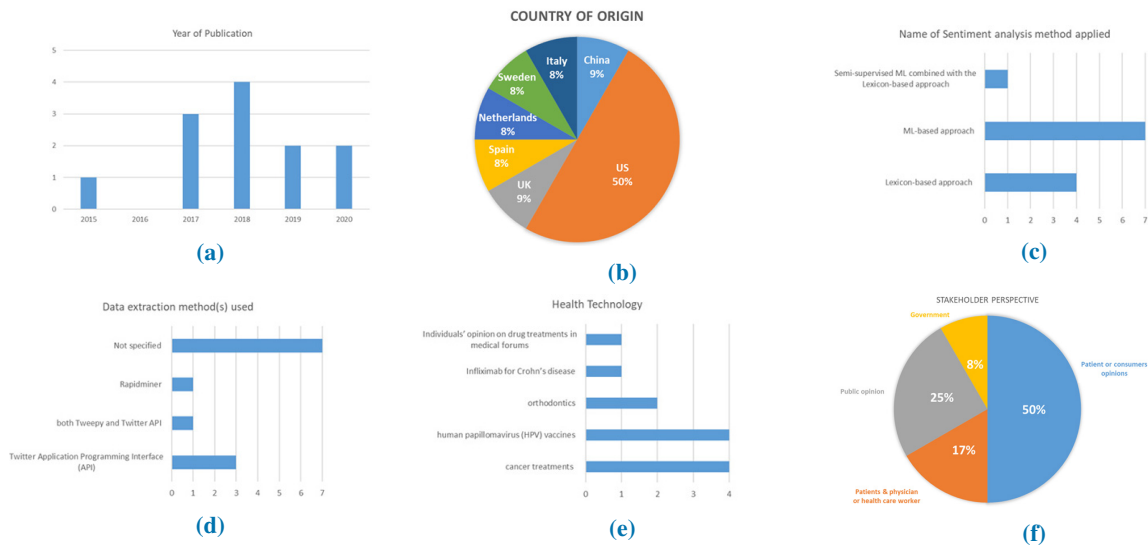
Cabling commented on the relevance of the SA of on-line support group messages to identify the topics being discussed, understand how users are talking about specific patient protocols, and how those that participate actively may engage in different topics than those who do not participate actively [21].

### c: Combination of Machine Learning-and Lexicon-Based Methods:

Opinions about physicians are easier to classify than opinions about the drugs prescribed by them. It was observed that the supervised learning method provided more accurate results than the lexicon-based approach alone [24].

Three studies did not discuss the strengths of the SA methods used but acknowledged social media as a valuable data source to better understand how stakeholders communicate their opinions about the available treatments received [16], [19], [17].

### 4) Limitations of Methods/Techniques as Described by the Authors



**FIGURE 2:** Graphical analysis of results. (a) The histogram presents the number of publications between 2015 and 2020. (b) The pie chart presents the country of publication of the selected studies. (c) The histogram illustrates the SA methods/techniques applied by the authors. (d) The histogram shows the data extraction method(s) used. (e) The histogram presents the health technologies assessed in the included studies. (f) The pie chart illustrates the proportion of stakeholder perspectives presented across the studies.

**a: Machine Learning-Based Method:**

Two significant drawbacks were noted in one study. The hierarchical Naïve Bayes classification technique requires a manual classification of a number of tweets to act as reference material to “train” the algorithm, and Twitter studies were unable to collect the demographic characteristics of users from their profile [22].

The corpus of tweets in one study was vastly imbalanced, so the distribution of different classes is highly diverse. As a result, it was difficult for ML-based methods to handle the classes with a limited number of tweets [23].

**b: Lexicon-Based Method:**

As the GCS technique is not customized to evaluate sentiments that show consumers’ resistance to or opinions about a medical product, the analysis sometimes misidentified the nature of the sentiment expressed. In many instances, an opinion about a health technology cannot be easily labeled into neat and distinct categories [18].

One study indicated that the overall sentiment tendency can be determined by TextBlob through calculating sentiment polarity scores based on lexicon, some relevant information (e.g. side effects) may be absent due to the 280 character limit on Twitter [19].

To reduce the time spent on reviewing the large volume of search results of YouTube videos, Livas recommended a more sophisticated screening approach that would identify relevant content through the suggested videos generated by the YouTube algorithm. It is uncertain if the proposed strategy represents the common practice in YouTube searches [25].

It was observed in one study that the patient sentiment on drug therapy (i.e., tamoxifen) may not be representative of the vast proportion of opinions as one user was responsible for 9% of posts, and 10 users were responsible for 30% of posts. These results align with other study findings on publicly accessible OSGs in which a marginal group of active users dominate most online forums [21].

Regardless of the automated tools used by the authors, they are unable to detect the nuances sometimes expressed in a human language (e.g., the context of the tweet and sarcasm). Hence, they may have underestimated the number of tweets as brand names were used to identify drugs, and tweets that used generic names or shortened versions were likely to have been missed.

**c: Combination of Machine Learning-and Lexicon-Based Methods:**

SA of public opinion on drug treatments that are characterized by the use of an informal language and lexical diversity can be a great challenge [24].

Limitations specific to the SA method used were not discussed in four studies [16], [20], [26],[17].

**IV. DISCUSSION**

Among the 12 selected studies in our scoping review, two studies examined patient experiences and opinions on orthodontic devices and invisalign treatments [25], [22], HPV vaccinations were the focus in four studies [18], [16], [23], [17] and the remaining studies targeted drug therapies. Even though our selection criteria encompassed a broad scope of health technologies, we were unable to identify relevant



studies on digital health technologies, such as mobile apps, for inclusion in our review. The ML-based approach had a greater representation in the review (seven studies), followed by four studies that used Lexicon based methods, and one in which a framework including both the approaches was proposed.

Each of the 12 studies used a different algorithm for opinion mining with different metrics and limitations. Of note, none of the selected studies in our review used a deep learning approach. This observation may suggest that, compared to deep learning approaches, researchers in the domain of health care technology find ML and Lexicon easier to use since these techniques are able to define specific parameters to support the opinion mining of different objects (e.g., treatments or medical devices). As a consequence, opinions about the same treatment or device that are investigated with different methods may be difficult to compare.

As a previous comparative analysis indicated [27], the agreement among the different SA tools may vary substantially, ranging from 33% to 80%. Furthermore, authors of the included studies identified a total of twenty limitations related to the use of SA (see: Supplementary File 2). These limitations can be categorized as follows: i) challenges in identifying and exporting relevant information due to the quality or quantity of the data available [19], [25]; ii) the need to adapt the parameters of SA to increase the accuracy of the analysis for specific contexts (e.g., supplement SA with human analysis to increase accuracy) [18], [24], [22]; iii) inability to extract data from jargon and informal communication (e.g., use of emojis to express feelings and opinions) [18], [24], [21], [26], [15]; and iv) challenges with the representativeness and groups dynamics [16], [21], [22].

The corpus of opinions across the social media platforms may not be representative of all patients who use the health technology in question. Groups dynamics [21], [22] may push people to be more active than others or been over-represented in the analysis, as well a digital divide may prevent relevant stakeholders to access the social network platforms [12]. SA, while enabling researchers to harvest data from all over the world, can also overestimate the representativeness of the data. SA can also be exposed to bias due to the misbehaviour of humans or artificial agents (e.g., trolling) which are used to nudge social networks discussions [28]. The phenomenon of trolling is often underestimated, while the ability of algorithms to discriminate noise from real insights of the target group is often overestimated [29]. Moreover, as Twitter and other social media have no access to demographic information since the user demographics are not linked to their profile, and social media use is not ubiquitous worldwide, SA studies will be limited to regions with high social media penetration [16].

Despite these limitations, the selected studies recognized the value of SA as a way to: i) account for emotional reactions [18], [16], [20], [15]; ii) rapidly and informally gather opinions of patients and other stakeholders from

different countries [18], [22], [23]; and iii) acquire new and unexplored perspectives regarding a topic [19], [24], [20], [21], [26] and explore less well-known issues related to the treatment or health technology [15]. SA can be also used as a way to include patients' opinions into healthcare decision-making processes [20], and it facilitates automatic aggregation and investigation of patients' decision-making behaviors, decision factors, social interaction trajectory decision-making.

The SA methods applied in the selected studies enabled the identification, classification, and analysis of data available on social media. The (semi)automatic quali-quantitative process of SA can help, for instance, public health experts to understand people reactions and investigate how to enhance their communication on HPV vaccinations, allow manufacturers to understand that there is a potential space for improving their products in a timelier manner, or to build a preliminary knowledge on certain topics that can be used by clinicians to facilitate patient-physician discussions on appropriate treatment options.

#### A. LIMITATIONS

While the literature search in our scoping review was limited to published studies, the main search concepts were deemed as broad by the authors. In addition, computer science databases were not part of our search strategy, but two major biomedical and health databases were searched to identify literature on the application of sentiment analysis to assess the safety and effectiveness of health technologies in clinical or "real-world" settings.

#### B. DIRECTIONS FOR FUTURE RESEARCH

One of the goals of our preliminary study is to design a method to analyze the performances of medical devices starting from data that are extracted from electronic medical records. A similar application can involve the analysis of technical reports after scheduled and corrective maintenance that are aimed to implement an evidence-based maintenance [30], [31], [32]. In addition, SA can play an important role in the post-market evaluation of medical devices as a feedback loop that consumers can use to express their satisfaction directly to the company [19]. SA can also precociously identify patients' needs and preferences and optimize the products and services that can lead to cost reduction and inform the development of personalized therapy plans. Future research can also investigate the development and application of a SA framework to extract free text data from social media platforms and generate RWE to support the development of HTA recommendations for decision makers. As such, the authors plan to design a method and tool for HTA, based on SA, framed in a three-year project of the International Federation of Medical and Biological Engineering / Health Technology Assessment Division (IFMBE/HTAD) [33].

## V. CONCLUSIONS

Our scoping review identified 12 studies on the use of SA, including ML-based and Lexicon methods, to assess public opinion on social media for specific health technologies. Two studies focused on medical devices, three examined HPV vaccinations, and the remaining studies targeted drug therapies. Due to the limitations and inherent differences among SA tools, the outcomes of these applications should be considered exploratory. The usefulness of SA lies on the quantity of data that can be rapidly collected and analysed to map the context and issues associated to a certain topic to preliminary inform further stages of systematic analysis which may inform decisions regarding treatments and devices. The results of our study can be an impetus for discussions on how the automation of algorithms developed to interpret public opinion of health technologies should be further developed to optimize the use of data available on social media.

## CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

## REFERENCES

- [1] HtaGlossary, "HtaGlossary.net | HomePage," 2020.
- [2] C. S. Goodman, HTA 101: introduction to health technology assessment. National Library of Medicine, 2014.
- [3] B. Schurman, "The Framework for FDA's Real-World Evidence Program," *Applied Clinical Trials*, vol. 28, no. 4, pp. 15–17, 2019.
- [4] S. Gohil, S. Vuik, and A. Darzi, "Sentiment analysis of health care tweets: review of the methods used," *JMIR public health and surveillance*, vol. 4, no. 2, p. e43, 2018.
- [5] O. Kolchyna, T. Souza, P. C. Treleven12, and T. Aste12, "Methodology for twitter sentiment analysis," arXiv preprint arXiv:1507.00955, 2015.
- [6] W. X. Zhao, J. Jiang, H. Yan, and X. Li, "Jointly modeling aspects and opinions with a maxent-lda hybrid," in *Proceedings of the 2010 conference on empirical methods in natural language processing*, pp. 56–65, Association for Computational Linguistics, 2010.
- [7] H. Kang, S. J. Yoo, and D. Han, "Senti-lexicon and improved naïve bayes algorithms for sentiment analysis of restaurant reviews," *Expert Systems with Applications*, vol. 39, no. 5, pp. 6000–6010, 2012.
- [8] M. Jain, S. Narayan, P. Balaji, A. Bhowmick, R. K. Muthu, et al., "Speech emotion recognition using support vector machine," arXiv preprint arXiv:2002.07590, 2020.
- [9] D. Kansara and V. Sawant, "Comparison of traditional machine learning and deep learning approaches for sentiment analysis," in *Advanced Computing Technologies and Applications*, pp. 365–377, Springer, 2020.
- [10] D. Levac, H. Colquhoun, and K. K. O'Brien, "Scoping studies: advancing the methodology," *Implementation science*, vol. 5, no. 1, p. 69, 2010.
- [11] A. C. Tricco, E. Lillie, W. Zarin, K. O'Brien, H. Colquhoun, M. Kastner, D. Levac, C. Ng, J. P. Sharpe, K. Wilson, and others, "A scoping review on the conduct and reporting of scoping reviews," *BMC medical research methodology*, vol. 16, no. 1, p. 15, 2016.
- [12] A. C. Tricco, E. Lillie, W. Zarin, K. K. O'Brien, H. Colquhoun, D. Levac, D. Moher, M. D. Peters, T. Horsley, L. Weeks, et al., "Prisma extension for scoping reviews (prisma-scr): checklist and explanation," *Annals of internal medicine*, vol. 169, no. 7, pp. 467–473, 2018.
- [13] M. D. J. Peters, C. M. Godfrey, H. Khalil, P. McInerney, D. Parker, and C. B. Soares, "Guidance for conducting systematic scoping reviews," *International journal of evidence-based healthcare*, vol. 13, no. 3, pp. 141–146, 2015.
- [14] H. Arksey and L. O'Malley, "Scoping studies: towards a methodological framework," *International journal of social research methodology*, vol. 8, no. 1, pp. 19–32, 2005.
- [15] A. Akay, A. Dragomir, and B. E. Erlandsson, "Network-based modeling and intelligent data mining of social media for improving care," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, pp. 210–218, 1 2015.
- [16] A. Budenz, A. Klassen, A. Leader, K. Fisher, E. Yom-Tov, and P. Massey, "Hpv vaccine, twitter, and gay, bisexual and other men who have sex with men," *Health Promotion International*, vol. 35, no. 2, pp. 290–300, 2020.
- [17] Q. Wang, W. Zhang, H. Cai, and Y. Cao, "Understanding the perceptions of chinese women of the commercially available domestic and imported hpv vaccine: A semantic network analysis," *Vaccine*, 2020.
- [18] X. Luo, G. Zimet, and S. Shah, "A natural language processing framework to analyse the opinions on HPV vaccination reflected in twitter over 10 years (2008 - 2017)," *Human Vaccines & Immunotherapeutics*, vol. 15, pp. 1496–1504, 8 2019.
- [19] L. Zhang, M. Hall, and D. Bastola, "Utilizing Twitter data for analysis of chemotherapy," *International Journal of Medical Informatics*, vol. 120, pp. 92–100, 12 2018.
- [20] D. De Silva, W. Ranasinghe, T. Bandaragoda, A. Adikari, N. Mills, L. Iddamalagoda, D. Alahakoon, N. Lawrentschuk, R. Persad, E. Osipov, et al., "Machine learning to support social media empowered patients in cancer care and cancer treatment decisions," *PloS one*, vol. 13, no. 10, p. e0205855, 2018.
- [21] M. L. Cabling, J. W. Turner, A. Hurtado-de Mendoza, Y. Zhang, X. Jiang, F. Drago, and V. B. Sheppard, "Sentiment Analysis of an Online Breast Cancer Support Group: Communicating about Tamoxifen," *Health Communication*, vol. 33, pp. 1158–1165, 9 2018.
- [22] D. Noll, B. Mahon, B. Shroff, C. Carrico, and S. J. Lindauer, "Twitter analysis of the orthodontic patient experience with braces vs Invisalign," *Angle Orthodontist*, vol. 87, pp. 377–383, 5 2017.
- [23] J. Du, J. Xu, H. Song, X. Liu, and C. Tao, "Optimization on machine learning based approaches for sentiment analysis on HPV vaccines related tweets," *Journal of Biomedical Semantics*, vol. 8, 3 2017.
- [24] S. M. Jiménez-Zafra, M. T. Martín-Valdivia, M. D. Molina-González, and L. A. Ureña-López, "How do we talk about doctors and drugs? Sentiment analysis in forums expressing opinions for medical domain," *Artificial Intelligence in Medicine*, vol. 93, pp. 50–57, 1 2019.
- [25] C. Livas, K. Delli, and N. Pandis, "'My Invisalign experience': content, metrics and comment sentiment analysis of the most popular patient testimonials on YouTube," *Progress in Orthodontics*, vol. 19, 12 2018.
- [26] M. Rocchetti, P. Salomoni, C. Prandi, G. Marfia, and S. Mirri, "On the interpretation of the effects of the Infleximab treatment on Crohn's disease patients from Facebook posts: a human vs. machine comparison," *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 6, 12 2017.
- [27] P. Gonçalves, M. Araújo, F. Benevenuto, and M. Cha, "Comparing and combining sentiment analysis methods," in *Proceedings of the first ACM conference on Online social networks*, pp. 27–38, 2013.
- [28] J. Paavola, T. Helo, H. Jalonen, M. Sartonen, and A. Huhtinen, "Understanding the trolling phenomenon: The automated detection of bots and cyborgs in the social media," *Journal of Information Warfare*, vol. 15, no. 4, pp. 100–111, 2016.
- [29] C. W. Seah, H. L. Chieu, K. M. A. Chai, L.-N. Teow, and L. W. Yeong, "Troll detection by domain-adapting sentiment analysis," in *2015 18th International Conference on Information Fusion (Fusion)*, pp. 792–799, IEEE, 2015.
- [30] V. Gonnelli, F. Satta, F. Frosini, and E. Iadanza, "Evidence-based approach to medical equipment maintenance monitoring," in *EMBECC & NBC 2017*, pp. 258–261, Springer, Singapore, 2017.
- [31] D. Medenou, L. A. Fagbemi, R. C. Houessouvo, T. R. Jossou, M. H. Ahouandjinou, D. Piaggio, C.-D. A. Kinnouezan, G. A. Monteiro, M. A. Idrissou, E. Iadanza, et al., "Medical devices in sub-saharan africa: optimal assistance via a computerized maintenance management system (cmms) in benin," *Health and Technology*, vol. 9, no. 3, pp. 219–232, 2019.
- [32] E. Iadanza, V. Gonnelli, F. Satta, and M. Gherardelli, "Evidence-based medical equipment management: a convenient implementation," *Medical & biological engineering & computing*, vol. 57, no. 10, pp. 2215–2230, 2019.
- [33] L. Pecchia, N. Pallikarakis, R. Magjarevic, and E. Iadanza, "Health technology assessment and biomedical engineering: Global trends, gaps and opportunities," *Medical engineering & physics*, vol. 72, pp. 19–26, 2019.

...