



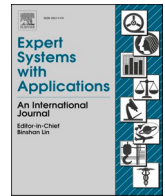
Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Contents lists available at ScienceDirect

Expert Systems With Applications

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

Review

Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: A systematic review

A.H. Alamoodi^{a,*}, B.B. Zaidan^{a,g}, A.A. Zaidan^a, O.S. Albahri^a, K.I. Mohammed^a, R.Q. Malik^b, E.M. Almahdi^a, M.A. Chyad^a, Z. Tareq^e, A.S. Albahri^f, Hamsa Hameed^c, Musaab Alaa^d^a Department of Computing, Sultan Idris University of Education (UPSİ), Tanjung Malim, Malaysia^b Department of Engineering Technology, Universiti Tun Hussein Onn (UTHM), Batu Pahat, Malaysia^c Faculty of Human Development, Sultan Idris University of Education (UPSİ), Tanjung Malim, Malaysia^d Faculty of Languages and Communication, Sultan Idris University of Education (UPSİ), Tanjung Malim, Malaysia^e Department of Computer Science, Computer Science and Mathematics College, Tikrit University, Tikrit 34001, Iraq^f Informatics Institute for Postgraduate Studies (IIPS), Iraqi Commission for Computers and Informatics (ICCI), Baghdad, Iraq^g Future Technology Research Center, National Yunlin University of Science and Technology, 123 University Road, Section 3, Douliou, Yunlin 64002, Taiwan, ROC

ARTICLE INFO

Keywords:

Sentiment analysis
COVID-19
Opinion mining
Disease mitigation
Epidemic
Pandemic
Infectious disease

ABSTRACT

The COVID-19 pandemic caused by the novel coronavirus SARS-CoV-2 occurred unexpectedly in China in December 2019. Tens of millions of confirmed cases and more than hundreds of thousands of confirmed deaths are reported worldwide according to the World Health Organisation. News about the virus is spreading all over social media websites. Consequently, these social media outlets are experiencing and presenting different views, opinions and emotions during various outbreak-related incidents. For computer scientists and researchers, big data are valuable assets for understanding people's sentiments regarding current events, especially those related to the pandemic. Therefore, analysing these sentiments will yield remarkable findings. To the best of our knowledge, previous related studies have focused on one kind of infectious disease. No previous study has examined multiple diseases via sentiment analysis. Accordingly, this research aimed to review and analyse articles about the occurrence of different types of infectious diseases, such as epidemics, pandemics, viruses or outbreaks, during the last 10 years, understand the application of sentiment analysis and obtain the most important literature findings. Articles on related topics were systematically searched in five major databases, namely, ScienceDirect, PubMed, Web of Science, IEEE Xplore and Scopus, from 1 January 2010 to 30 June 2020. These indices were considered sufficiently extensive and reliable to cover our scope of the literature. Articles were selected based on our inclusion and exclusion criteria for the systematic review, with a total of $n = 28$ articles selected. All these articles were formed into a coherent taxonomy to describe the corresponding current standpoints in the literature in accordance with four main categories: lexicon-based models, machine learning-based models, hybrid-based models and individuals. The obtained articles were categorised into motivations related to disease mitigation, data analysis and challenges faced by researchers with respect to data, social media platforms and community. Other aspects, such as the protocol being followed by the systematic review and demographic statistics of the literature distribution, were included in the review. Interesting patterns were observed in the literature, and the identified articles were grouped accordingly. This study emphasised the current standpoint and opportunities for research in this area and promoted additional efforts towards the understanding of this research field.

* Corresponding author.

E-mail addresses: Alamoodi@fskik.upsi.edu.my (A.H. Alamoodi), bilalbahaa@fskik.upsi.edu.my (B.B. Zaidan), aws.alaa@gmail.com (A.A. Zaidan), osamahsh89@gmail.com (O.S. Albahri), khalidibrahimm81@gmail.com (K.I. Mohammed), ramiqays@gmail.com (R.Q. Malik), esam.motasher@gmail.com (E.M. Almahdi), Mohammed.chyad1@gmail.com (M.A. Chyad), Ziadoontareq@tu.edu.iq (Z. Tareq), ahmed.bahri1978@iips.icci.edu.iq (A.S. Albahri), hamsahameed@fpm.upsi.edu.my (H. Hameed), mussabalaa@yahoo.com (M. Alaa).

<https://doi.org/10.1016/j.eswa.2020.114155>

Received 21 April 2020; Received in revised form 23 October 2020; Accepted 23 October 2020

Available online 28 October 2020

0957-4174/© 2020 Elsevier Ltd. All rights reserved.

1. Introduction

COVID-19 derived from *SARS-CoV-2* is currently spreading dramatically worldwide and causing millions of infections and deaths amongst the human population (Liu et al., 2020). *SARS-CoV-2* was detected in China in late 2019 in a general seafood marketplace and eventually infected millions of people (Velavan & Meyer, 2020). The situational reports of the World Health Organisation's (WHO) statistics have indicated that the number of confirmed cases exceeds 37,109,851, and the number of deaths exceeds 1,070,355 worldwide (WHO, 2020). Accurate insights into *COVID-19* can only be obtained when the pandemic ends as literature and statistics are proliferating, and keeping data updated is nearly impossible (Hamzah et al., 2020). On 28 February 2020, the WHO launched emergency protocols in all medical and public health systems because of the severity and risks of *COVID-19* (Epidemiol., 2020). *COVID-19* is not the first global pandemic. Several different viruses and pandemics, including Ebola (McMullan, 2020), Mers-Cov and SARS, have occurred in the past. Medical doctors and medical researchers have earnestly dealt with these pandemics, and their efforts have not been in vain (Elder, Johnston, Wallis, & Crilly, 2020). Nevertheless, with the current trends of technologies, especially the role of computer science, computer technologies have fairly shown their contribution to medical decisions, such as infectious diseases and outbreaks (Bhat et al., 2020; Soliman, Tabak, & Sciences, 2020). Historical data are utilised in the process, and an increase in the availability of data enables researchers to generate better decisions and conclusions (Pan et al., 2020). Current and genuinely reasonable sources for obtaining these data include social media platforms that provide available data more than ever before. Interestingly, these data serve as the basis for conducting opinion mining and sentiment analysis.

1.1. Social media and sentiment analysis

Social media has launched platforms to facilitate communications amongst human communities and help them share ideas, information, knowledge and other data by these forms of electronic communication (K.-S. Kim, Sin, & Yoo-Lee, 2014). Social media platforms are gaining remarkable influence than ever before and are considered one of the fastest growing information systems for social applications (Appel, Grewal, Hadi, & Stephen, 2020; Xu, Zayed, Lin, Wang, & Li, 2020). Social media platforms are considered the global centre of big data as people use their applications and spend excessive hours on these media outlets (DeNardis & Hackl, 2015). Some of the most commonly employed social media applications in the world are Facebook, Twitter, Instagram and Reddit (K. Ali, Dong, Bouguettaya, Erradi, & Hadjidj, 2017). Social and statistical studies have shown that these applications influence human behaviours, given the users' length of time spent on them, which ranges from hours per week to daily use (Statista, 2019). Despite the large data presented on these social media platforms, their content may have contradictory effects, which range from negative psychological influence on people's lives to positive psychological influence on people's lives (Crawford, 2009). People who are addicted to social media likely unleash and share opinions and ideas across these platforms (Jansen, Sobel, & Cook, 2010). Subsequently, turning these opinions and posts into assets is highly valuable. Discovering a Tweet or Facebook post may be possible with millions of likes and retweets, but this massive interaction with such a post does not reflect its importance or the emotions of users who participate in the post because of many factors, such as the nature of posts, including negation and irony (Ji, Chun, Wei, Geller, & Mining, 2015); happiness and sadness (K. Ali et al., 2017); anger (Ji et al., 2015); positive and negative (Zarrad, Jaloud, & Alsmadi, 2014); concern, surprise, disgust or confusion (Ji, Chun, & Geller, 2016); and the massive numbers of tweets (Gayo-Avello et al., 2013). Nevertheless, large-scale extractions of human emotions and entertainment from social media networks are essential for international public influences, business decisions and policy development (Chung,

He, & Zeng, 2015). Sentiment analysis and opinion mining have become useful. Sentiment analysis contributes to the understanding of human emotions as it can seek people's behaviours as users engage in these social media applications (Ji et al., 2016). Additionally, these media applications have been employed in various application domains, including tourism (Ainin, Feizollah, Anuar, & Abdullah, 2020), business (Reyes-Menendez, Saura, & Filipe, 2020), education (Hassan et al., 2020) and health (Rodrigues, das Dores, Camilo-Junior, & Rosa, 2016), for various beneficial purposes, such as analysing opinions (Zarrad et al., 2014) and allowing people to express their emotions freely (Chung et al., 2015), and for highly dynamic and real-time data trends (Chaudhary & Naaz, 2017). With this feature, large-scale communities can be observed at a low cost (Choi et al., 2017). Therefore, sentiment analysis is a powerful tool in understanding the most important events and trends.

1.2. Sentiment analysis during epidemics and disease outbreaks

Computer technologies provide profound opportunities to fight infectious disease outbreaks (Eysenbach, 2003; Goldschmidt, 2020) and have a remarkable role, especially in sentiment analysis for social media (Singh, Singh, & Bhatia, 2018); this importance is due to their tremendous role in analysing public sentiment. Various research articles have indicated that many outbreaks and pandemics could have been promptly controlled if experts considered social media data (Singh et al., 2018). Therefore, sentiment analyses in studying pandemics, such as *COVID-19*, are important based on recent events. *COVID-19* remains a controversial global topic in social media (Pastor, 2020). This review aimed to examine the role of sentiment analysis in the occurrence of *COVID-19* and other previous infectious diseases via a systematic review protocol that involved previous research-related efforts adopted for a span of 10 years.

1.3. How does sentiment analysis work

This section illustrates the process of sentiment analysis. Although some authors have taken different steps, Fig. 1 shows the main flow of sentiment analysis.

- *First Step: Social Media Platforms:* In this phase, a user selects the source from where he/she wishes to extract a sentiment; for example, in social media selection, a user can choose various online sources, such as Facebook, Twitter and Reddit (Chung et al., 2015).
- *Second Step: Data Collection Process:* In this phase, users start with the use of certain keywords (Deng, Tang, & Huang, 2015) or hash tags (*i.e.*, #) (Gayo-Avello et al., 2013) to acquire their desired information based on their preferences. This information has different forms (*e.g.*, tweets, posts, news and texts) (Chung et al., 2015).
- *Third Step: Preprocessing:* In this phase, the extracted information is processed to prepare the data for the next phase (Kim, Jeong, Kim, Kang, & Song, 2016). This stage includes feature extraction (Jain & Kumar, 2018) (*i.e.*, grammatical structures and mining characteristics), tokenisation (Baker, Shatnawi, Rawashdeh, Al-Smadi, & Jararweh, 2020) (*i.e.*, process of converting text into tokens before transforming it into vectors) and cleaning (E. H.-J. Kim et al., 2016; Pollacci et al., 2017; Zarrad et al., 2014) (*i.e.*, repeated letter removal, text correction, normalisation, stop word removal and language detection).
- *Fourth Step: Data Analysis:* In this phase, all preprocessed data are utilised for their intended purposes, such as polarity identification (E. H.-J. Kim et al., 2016), sentiment analysis (E. H.-J. Kim et al., 2016) or frequency analysis (Deng et al., 2015).

2. Systematic review protocol

This study was conducted based on a systematic review protocol (SLR) that helps achieve a comprehensive understanding of the research interest while providing further information for future studies

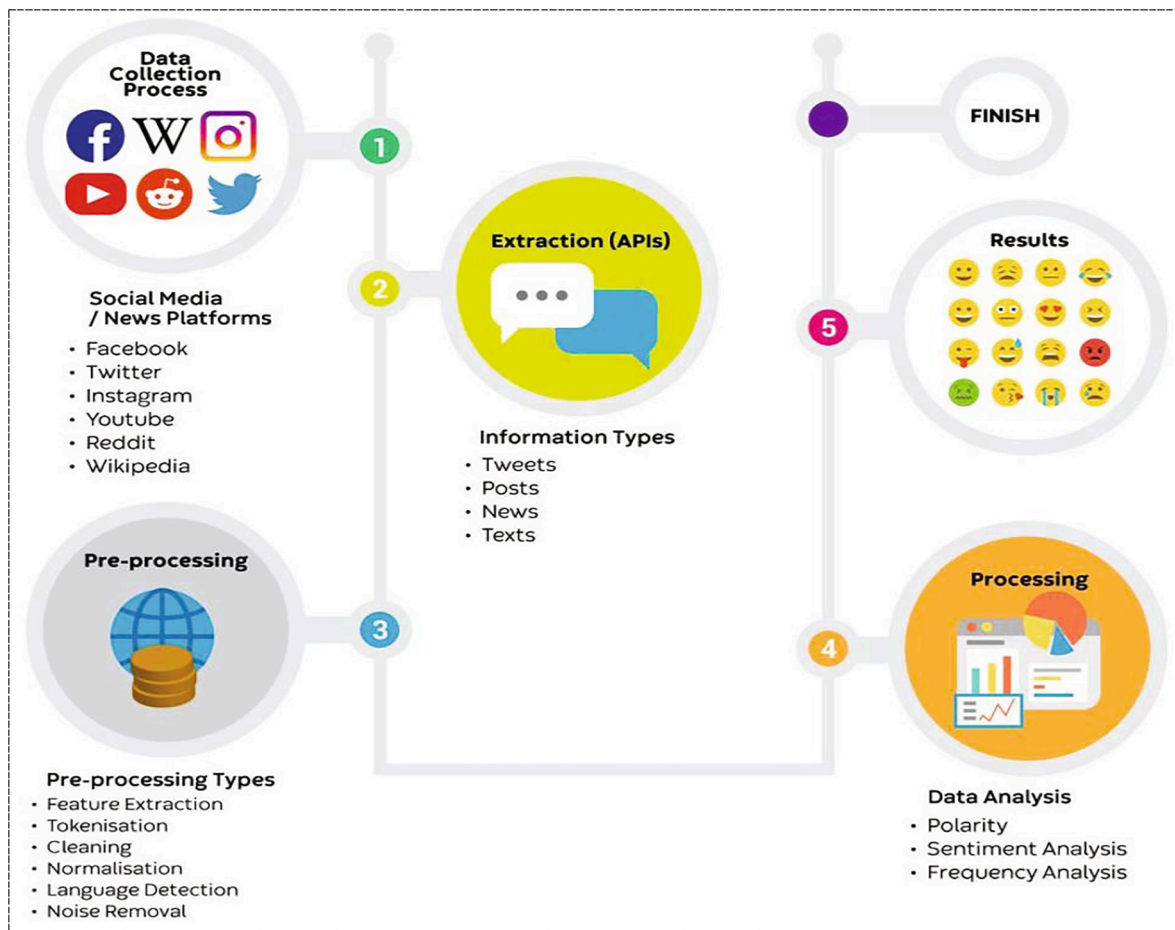


Fig. 1. Sentiment analysis flow.

(Alamoodi et al., 2020). In addition, this research approach is considered a more well-structured method for research synthesis than conventional approaches because of its methodological process and identification metrics in identifying relevant studies (Burgers, Brugman, & Boeynaems, 2019; Kushwah, Dhir, Sagar, & Gupta, 2019; Loureiro, Romero, & Bilro, 2019). Moreover, the systematic review approach was identified as an advanced method because of its large impact on various research areas and scientific disciplines. The systematic review principally consisted of a set of processes, such as research area identification, searching technique, study selection and information extraction and synthesis (Ain, Vaia, DeLone, & Waheed, 2019). In the systematic review, the term ‘primary studies’ was given to individual studies, in which original data were collected and analysed, whereas the term ‘secondary studies’ was assigned to other studies (i.e., studies with data collected and analysed previously) (Alamoodi et al., 2019).

2.1. Information source

The applied search strategy in this review was based on preferred reporting items for systematic review and meta-analysis (PRISMA) statements (Alamoodi et al., 2019), as summarised in Fig. 2. Five scientific digital databases were chosen for the article search, download, filtrations, extractions and drafting of this review: (1) Science Direct database, which offers large access to scientific papers across different academic disciplines; (2) Scopus database, which provides various publication contents that are related to different and various domains of science; (3) IEEE Xplore, which has scientific soundness and multidisciplinary technology publications on different domains; (4) PubMed, which features different publications from medical and technological

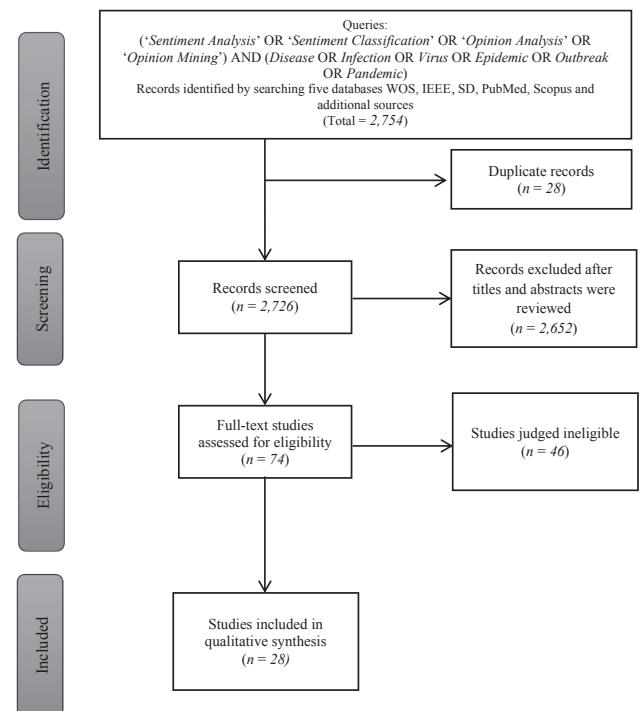


Fig. 2. SLR protocol.

sciences; and (5) Web of Science, which has a wide range of publications across different domains, such as social sciences, arts and humanities. The selected databases showed tremendous occurrences in many of the published systematic reviews on high-impact scientific journals, which have academic resilience and scientific soundness (Diedrichs & Services, 2000; Falagas, Pitsouni, Malietzis, & Pappas, 2008; Gies, 2018; Griffin, 2002; Meho, Yang, & technology, 2007; Tober, 2011). Thus, these databases were deemed sufficient and most suitable for this review. Additional search resources were considered for their significance in text analytics and employed in this review.

2.2. Search strategy

The search was carried out in two stages: the first stage was conducted on 31 March 2020 after the drafting of the main highlights of the manuscript, and the second stage was carried out on 30 June 2020 to ensure that more updated and recent literature is included. The search was initiated in the advanced search boxes of the previously mentioned scientific databases. Conjunctive and disjunctive models with Boolean operators were utilised for the search (*i.e.*, AND, OR), and two groups of keywords (*i.e.*, queries) were utilised in the process, as presented in Fig. 2. The previous process was performed to retrieve the most related articles. Despite the availability of other retrieving models, such as cosine similarity, which is considered a measure that has been extensively applied in pattern recognition and text classification to measure how the documents are similar (Al-Anzi & AbuZeina, 2017; Aljuaid, Iftikhar, Ahmad, Asif, & Afzal, 2020), or other semantic similarity models (Rotaru & Vigliocco, 2020), these models exhibit some limitations, such as data access and inability to apply some search filters (*i.e.*, publication types, languages, specific data search or search areas). Thus, the selection of these models was not the best practice in conducting a systematic review. Note that cosine similarities and other semantic similarity models are most suited for distance measures in text classification problems (Al-Anzi & AbuZeina, 2017), while Boolean operators were chosen here due to their data source and access capabilities. Therefore, choosing Boolean operators was more practical, and thus, utilised in this review. In searching and filtration, content on the basis of various types of publication articles, conferences, books and review articles was chosen. This option was deemed efficient for covering the latest and most relevant publications in the designated topic of this review.

2.3. Study selection

This research process was initiated and composed of three sub-processes, namely, article collection, title and abstract scanning and full-text reading (Elaish, Shuib, Ghani, Yadegaridehkordi, & Alaa, 2017; Liberati et al., 2009). In the first process, the articles with the initial number $n = 2754$ were collected from the selected databases, and $n = 28$ duplicate articles were scanned across all the databases. In the next process, relevant articles were determined by scanning their abstracts and titles. All the extracted articles were thoroughly investigated by reading their titles and abstracts to verify whether they met the inclusion criteria discussed in Section 2.5. If an article was a match, then it was included in the final round. In the third process, the entire full-text reading was performed. A study was excluded if it did not fit the criteria of this review; a total of $n = 2,652$ unrelated papers were identified. Full-text reading was conducted, and useful and valuable information (*i.e.*, data extracted) was extracted for the final set of articles that satisfied the inclusion criteria identified in this study $n = 28$. Several authors' notes and comments (*i.e.*, data extraction element drafting) were collected during the process, as shown in Table 1, which evolved to provide insights that shaped the final form of this review. All these details and the corresponding processes are discussed in Section 2.4.

Table 1
Data extraction elements.

Data Item	Description
Title	(Title of the paper)
Year	(Publication Year)
Database	Name of the database where an article was downloaded (<i>i.e.</i> Science Direct/ Scopus/Web of Science/IEEE and PubMed)
Publication Type	Journal/Conference/Book
Outbreak Discussed	Name of (<i>i.e.</i> virus, pandemic, infectious disease, outbreak)
Goal	Primary goal of the study (<i>aim of the study</i>)
Source Data	Where the data used in the sentiment analysis were collected from (<i>i.e.</i> Twitter, Facebook, Reddit, etc.)
Volume of Data	Size of data (How many Tweets, how many posts, etc.)
Duration of Collection	How long the data collection lasted for sentiment analysis (<i>i.e.</i> Days, Months, Years)
Challenges	Issues or concerns raised in the publication
Motivations	Significance or benefits identified in the publication

2.4. Data extraction

Data were collected and extracted to analyse every article for various attributes, which were subsequently listed and grouped in matching categories by using an Excel spreadsheet. These attributes, as shown in Table 1, were considered the most important aspects to obtain the main discussion points of this systematic review. This procedure was followed by a summary and table description of the main findings.

Each of these data extraction elements was chosen for a reason. The first reason was related to the article title and meant to show the title of the paper for future referencing and for any reader who might be interested in perusing the article. The publication year, database selected, publication type and name of the outbreak that was discussed were extracted to show the readers the demographic statistics of this review, which showed how these publication types progressed from 2010 to 2020 by presenting the increase in the number of studies. The readers could determine which database held the most papers on the related topic, so it could be considered when other investigations on related topics would be designed for future research. The publication type was meant to show the readers which journals were interested in these topics by providing them publishing sources for potential future work, which was related to infectious diseases and sentiment analysis. This review could also present the conferences related to sentiment analysis. The last component of the demographic statistics related to the outbreak discussed was extracted to reveal previously discussed topics and present them as references for future research on these topics. Four data elements, namely, goal, source data, data volume and collection duration, were extracted to design the discussion and categorisation of taxonomy, with the table that discussed the details of the taxonomy analysis. The latter showed how articles were categorised across different domains, which data source was previously considered when someone worked on similar topics and which volume of data and durations of their collection were involved in previous studies to guide future studies. The last two elements were designed to convey the message of previous researchers to future peers. Presenting the challenges was one way of informing new researchers about the problems encountered by their peers and providing a new topic for future research. Detailing the motivations was the previous author's process of demonstrating to their future peers the significance and advantages of working on these topics.

2.5. Inclusion and exclusion criteria

Several inclusion and exclusion criteria were imposed while attempting to identify the most relevant articles in the study selection process. The date of publication was set from 2010 to 30 June 2020. In accordance with additional criteria, all papers, which included reviews, conferences, books and research papers, were limited to those in the

English language across all the selected databases. Additional search resources were considered for their significance in text analytics: conference proceedings from the International Workshop on Semantic Evaluation (*SemEval*), Association for Computational Linguistics (*ACL*), North American Chapter of the Association for Computational Linguistics (*NAACL*), Cross-Language Evaluation Forum (*CLEF*) and PAN series of scientific events and shared tasks on digital text forensics and stylometry. The remaining criteria were concerned with the inclusion of all the papers that discuss the health aspects of sentiment analysis and opinion mining, whose focus was a discussion of infectious diseases or outbreaks, such as epidemic, pandemic or virus spread, as shown in [Table 2](#).

In accordance with some criteria, papers such as those not written in English or older than 2010 were excluded. The remaining exclusion criteria were studies that discussed sentiment analysis with normal diseases (*i.e., non-infectious*), such as cancer. Articles that discussed infectious diseases without sentiment analysis were also excluded as they were not within the scope of this review and exclusively related to medicine and health.

2.6. Demographic statistics

This section describes some of the key demographic statistical findings from the research results in [Fig. 3](#).

[Fig. 3](#) shows different statistical values from the selected articles. The upper-left figure describes the article distribution in terms of the publication year. On the bottom-left side, the article resources of the included papers are presented with the overall occurrence in each database. In the top right of the figure, the topics and names of infectious diseases and the number of occurrences across the publications included in this review are presented. In the last part on the bottom right of the figure, the publication types of the selected articles are described.

3. Taxonomy

This section presents our taxonomy, which summarises the results of our search process. First, the articles were searched, scanned and filtered. Second, full-text reading of all selected articles $n = 28$ was performed. Afterwards, all the articles were classified into four major categories. These major categories were linked to the corresponding disease, infection or epidemic from the references. In the first category, lexicon-based model sentiment was discussed. In the second category,

Table 2
Inclusion and exclusion criteria.

	Inclusion Criteria	Exclusion Criteria		
Search	English Articles	Non-English Articles		
	Selected Databases			
	Article, Review, Conference or Book Section	Older than 2010		
	Published between 2010 and 2020			
Topic	Infectious Disease	AND	Sentiment Analysis OR Opinion Mining	Other diseases (<i>not infectious</i>)
	Pandemic			Without Sentiment Analysis Or Opinion Mining
	Epidemic			
	Outbreak			
	Virus			

machine learning (ML)-based sentiment models were addressed. In the third category, the hybrid literature between the previous two categories was described. In the fourth category, a section titled ‘individuals’ was included to cover the remaining articles that did not fit into the first three categories. These categories were classified based on a previous work by [Pollacci et al. \(2017\)](#), which required the sentiment analysis literature to be consistent with these categories. In addition, creating the taxonomy in this way enabled the categorisations of literature based on a common theme inspired by the reference and agreed upon during authors’ discussions. The articles in the taxonomy were presented using their appearance number in accordance with the taxonomy discussion. Each discussed disease was also presented with its number of articles from all the categories in [Fig. 4](#). All the taxonomy analyses that contain the most important data, including the reference, data source, volume of data, duration of collection, discussed outbreak and work applied, are presented in [Table 3](#).

3.1. Lexicon-based models

This category presented the first set of studies in the taxonomy that discussed the lexicon-based technique and their application in infectious diseases and sentiment analysis. This technique depended on the polarity score of the given text in accordance with their positive and negative values, which were addressed by the word dictionaries. The section contained $n = 10/28$ articles. In the first study by [E. H.-J. Kim et al. \(2016\)](#), the implications of multiple media sources, such as news in journal reports and tweets on Ebola disease, were discussed by acquiring Twitter data between 1 June 2014 and 31 August 2014. Their experimental results indicated the narrower and more blurry topic coverage for Twitter than the news media, with the latter also having a longer life span for sentiment dynamics than the former. In the second study by [Chung et al. \(2015\)](#), 255,118 tweets posted by 210,900 users in January 2015 were used to address the significance of the Ebola disease, and a medical health informatics system that is referred to as ‘eMood’ was developed to collect social media data and visualise the results of the collected data analysis. Two factors, namely, users’ centrality and influence, were investigated in this study. The results indicated the importance of social media understanding and the large impact on public health medical informatics organisations. In the third study by [Deng et al. \(2015\)](#), social media data on Ebola and the typhoons Haiyan and Hagupit were analysed using the distributed system approach for data mining and the dataset acquired in China from June 2014 to November 2014. The results indicated the importance of integrating geographical information with social media data, which could help coordinate strategies to mitigate disasters and side effects. Conversely, in the fourth study by [Zarrad et al. \(2014\)](#), the effect of coronavirus infection (*MERS-CoV*) in Saudi Arabia was discussed to address the existing challenges in big data platforms, such as ‘Apache Hadoop’, and analyse people’s opinions by extracting 1,500,000 tweets, which were collected within 3 months. The study strongly recommended the development of an automated system for people’s opinion analysis based on its outcomes to help decision makers in different governments and public sectors. In [K. Ali et al. \(2017\)](#), a new framework for sentiment analysis was developed by utilising the dataset collected from four social media platforms, namely, Instagram, news forums, Reddit and Twitter. The acquired data consisted of 525 tweets and posts. Furthermore, a new sentiment analysis model based on spatiotemporal properties was explored to localise the outbreaks of disease infections. In another study in this category by [Almazidy, Althani, and Mohammed \(2016\)](#), smart home dashboards were designed for social media data tracking using a data mining technique to notify users regarding disease outbreaks and locations based on Twitter data by using different application programming interfaces (*APIs*). This approach was recommended and considered an effective method for human–computer interaction (*HCI*). In the following work by [Bhat et al. \(2020\)](#), more than 80,000 tweets associated with the spread of *COVID-19* were subjected to sentiment

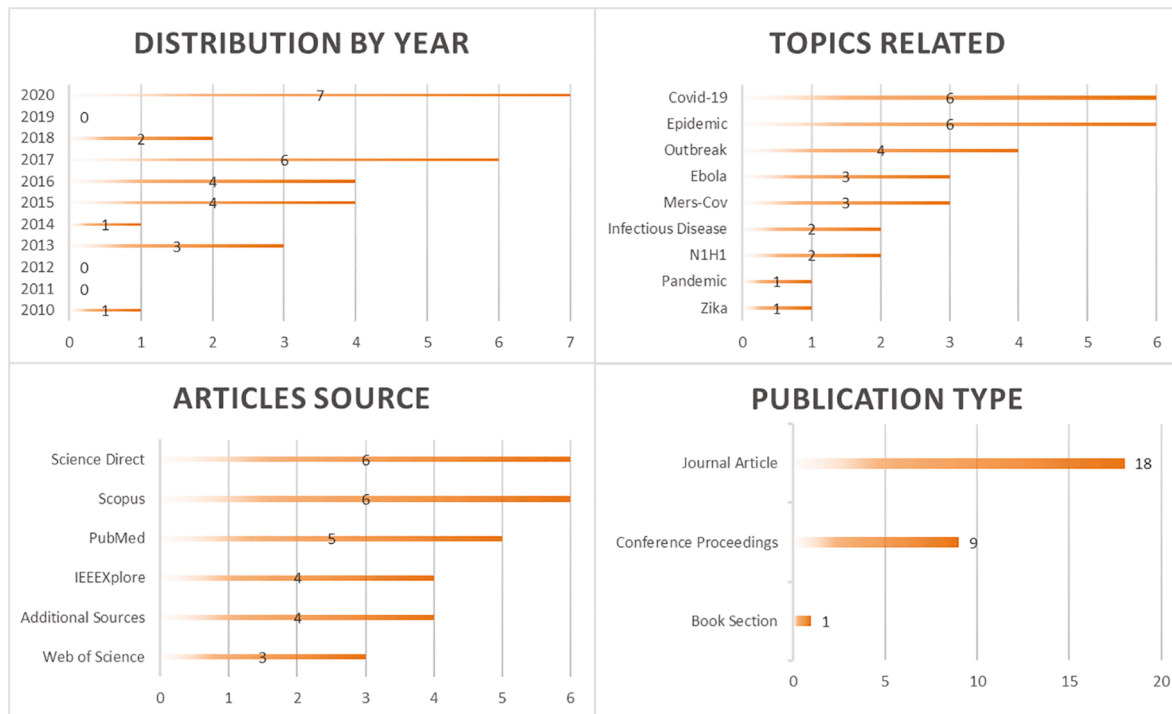


Fig. 3. Demographic statistics.

analysis. In particular, 51.97%, 34.05% and 13.96% of the tweets expressed positive sentiments, neutral sentiments and negative sentiments, respectively. The role of recent health technologies, such as artificial intelligence (AI), was described to contribute to future sentiment analysis for mitigating pandemics and assisting in various strategies, such as lockdown, social distancing and other strategies. In another work by Lwin et al. (2020), COVID-19 was discussed by examining the worldwide Twitter trends about the pandemic, utilising more than 20,325,929 tweets for the public emotion responses during the pandemic and applying a lexical approach and ‘CrystalFeel’ algorithm. Four emotions (fear, anger, sadness and joy) and the narratives that underlie these emotions were investigated. Negative emotions were dominant during the COVID-19 pandemic, which supports the recent call for action to maintain the public’s mental well-being during this unprecedented crisis. These emotions must be addressed and counter-balanced by complementing regular crisis communication with strategic public health communication. Another interesting work on the fight against COVID-19 by Raamkumar, Tan, and Wee (2020) aimed to understand the communication strategies that are conducted by various public health authorities to examine public sentiment and responses to COVID-19 in social media outlets. The main objective of the study was to clearly improve public information dissemination practices. Unlike Twitter posts, Facebook posts and comments were utilised, and sentiment analysis was measured by using frequency, mean sentiment polarity (SP), positive-to-negative sentiment ratio (PNSR) and positive-to-negative emotion ratio (PNER). Another study by Pastor (2020) investigated the sentiments of the Filipinos and the effect of extreme community quarantine caused by the COVID-19 pandemic via sentiment analysis on Twitter. Negative sentiments were more dominant than positive sentiments and expected to increase based on the current events as they affect personal lifestyle. This study recommended that continuous studies should be conducted to determine the sentiments of constituents in the Philippines, and these data would provide intervention or support in accordance with users’ sentiments. All previous papers discussed the related lexicon-based literature from the taxonomy. Fig. 4 also illustrates the usage of sentiment analysis in relation to Ebola $n = 3$, MERS-CoV $n = 1$, disease outbreaks $n = 2$ and COVID-19 $n = 4$.

3.2. Machine learning-based models

This category is the second taxonomy-derived type that discusses the applications of sentiment analysis to examine social media data on infectious diseases. Fundamentally, ML-based categorisation was employed for analysis, and based on the taxonomy of this study, only five studies were conducted on this type of research $n = 5/28$. In the first study by Lim, Tucker, and Kumara (2017), a model on Twitter data acquired from August 2012 to May 2013 was validated, and unsupervised learning techniques were applied because of their efficiency in the mining of multimedia data. The study aimed to detect the infectious disease by multimedia data mining and discovered that most of the existing systems involved the use of a down approach and relied on pre-existing knowledge, such as the name of a symptom and the name of a disease. A promising result was obtained after Twitter data were analysed, and the precision and recall as evaluation metrics were utilised. In Baker et al. (2020), Arabic tweets were explored to detect an epidemic. Influenza-related tweets that are written in Arabic were collected, labelled, filtered and analysed. A total of 54,065 Arabic tweets and locations were collected with the Twitter API and subsequently classified using four classifiers (i.e., decision trees, k-nearest neighbour, support vector machines and naive Bayes). The results of the classification revealed excellent potential from experiments, where naive Bayes and K-nearest neighbour achieved an accuracy of 89.06% and 86.43%, respectively. In another study by Culotta (2010), various classification methods were utilised to detect influenza-related messages (N1H1) from Twitter and compare the results with official statistics from the Centers for Disease Control and Prevention (CDC) by using more than 574,643 tweets. The best model acquired 84.29% accuracy by multiple linear regression, and the results suggested that the existing correlation with those of the CDC of 0.78 can be obtained by leveraging a document classifier to identify relevant messages. In another study by A. Ali, Magdy, and Vogel (2013), healthcare tweet monitoring platforms in different regions were presented by utilising an ML approach with a support vector machine (SVM) to automatically detect and classify healthcare tweets associated with MERS-CoV. This approach offers governments with a solution to monitor public health and measure public satisfaction with respect to healthcare

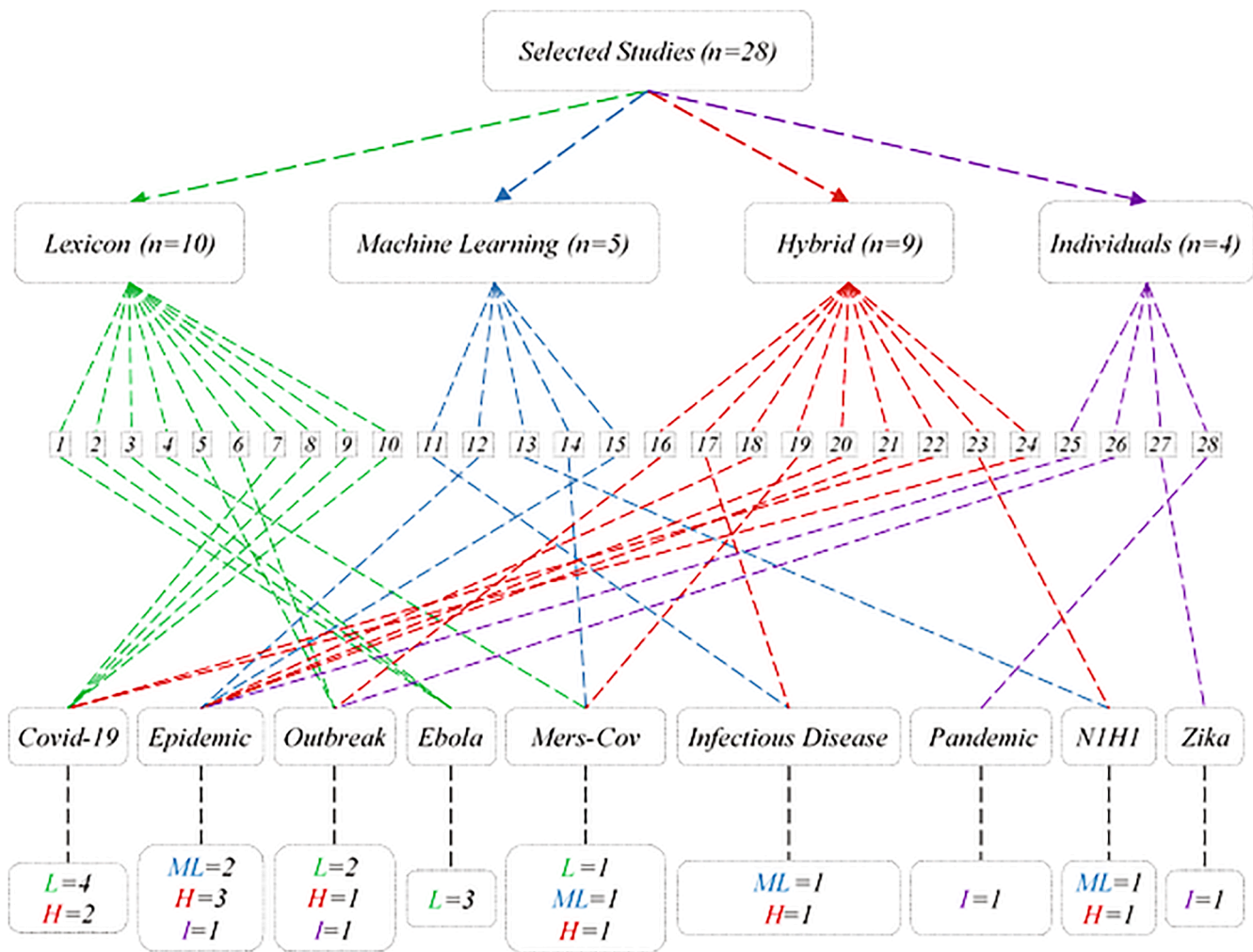


Fig. 4. Taxonomy for related literature.

services from people’s posts on Twitter. The use of the SVM in this regard was associated with the identification of tweets despite their relevance to healthcare. Another work by Singh et al. (2018) investigated some of the previous works on conducting ML-based sentiment analysis to predict epidemics and cover aspects such as the relevance of sentiment analysis via ML techniques and classification of data based on ML techniques. In addition, the review was associated with literature that shows the strength and drawbacks in the predictions of healthcare information and epidemics via ML with data sets available on social media, RSS news feeds and data from healthcare agencies. However, detailed discussions via sentiment analysis of other infectious disease aspects, such as pandemics and viruses, and the role of sentiment analysis in people’s emotions on Twitter during these events were excluded. The five papers discussed ML-related literature with sentiment analysis from the taxonomy discussion in Fig. 4. The usage of sentiment analysis in relation to infectious diseases $n = 1$, epidemic $n = 2$, N1H1 $n = 1$ and MERS-CoV $n = 1$ was also discussed.

3.3. Hybrid-based models

The third main section of the taxonomy discussed the literature on sentiment analysis with disease infections from both aspects of lexicon-based models and ML models (i.e., hybrid). The section contains a total of $n = 9/28$ articles. Apart from Ebola and previous outbreaks, mosquito-borne and outbreak management and disease surveillance were discussed based on social media by using spatial and temporal information, which facilitates the identification, characterisation and modelling of

user behavioural patterns on the web via sentiment analysis, as discussed in Jain and Kumar (2018). Two experiments were carried out using some of the most well-known lexicons (i.e., SentiwordNet and AFINN), and the tweets were classified using ML techniques, namely, SVM and NB. A predictive mapping based on geotagging data was proposed for use in a specific area with limited resources that enabled the tracking of real-time public sentiments for early discovery or alarm mechanisms related to outbreaks. Another work in Ji et al. (2015) discussed the issues of spreading public concern about infectious diseases by measuring public concern by using a two-step sentiment classification approach to keep track of Twitter trends. This approach comprises two steps. In the first step, the tweets that carry personal opinions of tweeters are separated from third-party factual reports, such as news articles with the use of clue-based lexicons. In the second step, two different ML classifiers, NB and SVM, were trained and tested for the data collected from 12 datasets of six infectious diseases, namely, listeria, influenza, swine flu, measles, meningitis and tuberculosis, with approximately 15 million tweets. In another study by Pollacci et al. (2017), an epidemic spreading-based approach was proposed to enhance lexicon-based sentiment analysis. In this approach, the lexicon automatically extends from a reduced dictionary and large amounts of Twitter data with 3,718 publicly available tweets and sentiments. For validation purposes, the ML technique, namely, SVM, was utilised with lexicon-based features to provide the best performance in most cases. Accordingly, the number of tweets labelled with the new dictionary increased by nearly 45%. In Choi et al. (2017), the lexicon approach and ML were employed by proposing a computational method to monitor

Table 3
Taxonomy analysis details.

Ref	Source Data	Volume of Data	Duration of Collection	Outbreak Discussed	Work Applied	Taxonomy Category
E. H.-J. Kim et al. (2016)	Twitter News Publications	16,189 news articles 7,106,297 tweets	Between 1 June 2014 and 31 August 2014	Ebola	Sentiment Dynamics of the Hot Health Issue of Ebola	Lexicon Based Models
Chung et al. (2015)	Twitter	255,118 tweets	January 2015	Ebola	Ebola Outbreak Discussions on Twitter	Lexicon Based Models
Deng et al. (2015)	Sina Tencent Sohu Netease Twitter	140,000 tweets 180,000 records	June to November 2014	Ebola Typhoon Haiyan Hagupit	Distributed Mining System for Online Opinion Data Collecting and Mining.	Lexicon Based Models
Zarrad et al. (2014)	Twitter	1,500,000 tweets	3 months	MERS-CoV	Addressing New Challenges for Big Data Platform with an Opinion Mining Approach	Lexicon Based Models
K. Ali et al. (2017)	Twitter Reddit Instagram News Fora	525 Reviews	2 Months	Outbreaks	Locations Monitoring for Disease Outbreak	Lexicon Based Models
Almazidy, Althani, and Mohammed (2016)	Twitter	N/A	N/A	Outbreak	Outbreak Notification Framework Using Twitter Mining for Smart Home Dashboards	Lexicon Based Models
Bhat et al. (2020)	Twitter	N/A	N/A	COVID-19	Sentiment analysis of Social Media Response on the COVID-19 Outbreak	Lexicon Based Models
Lwin et al. (2020)	Twitter	20,325,929 tweets	28 January 2020 to 9 April 2020	COVID-19	Examining Worldwide Trends of Different Emotions During the COVID-19 Pandemic.	Lexicon Based Models
Raamkumar, Tan, and Wee (2020)	Facebook	3,185,460 posts	1 January 2019 to 18 March 2020	COVID-19	Understand the Communication Strategies by Public Health Authorities to Examine Public Sentiment and Responses in Social Media	Lexicon Based Models
Pastor (2020)	Twitter	N/A	3 Weeks	COVID-19	Sentiment analysis of Filipinos and Effects of Extreme Community Quarantine due to the Coronavirus (COVID-19) Pandemic	Lexicon Based Models
Lim, Tucker, and Kumara (2017)	Twitter	37,599 Tweets	From August 2012 to May 2013	Infectious Diseases	Identify Real-World Latent Infectious Diseases by Mining Social Media Data	ML Based Models
Baker et al. (2020)	Twitter	54,065 Tweets	N/A	Epidemic	Detection Of Epidemic Disease Model for Discovering Latent Infectious Diseases	ML Based Models
Culotta (2010)	Twitter	574,643 Tweets	10 weeks	N1H1	Classification Methods for the Detection of Influenza-Related Messages (N1H1) from Twitter	ML Based Models
A. Ali, Magdy, and Vogel (2013)	Twitter	20,000 Tweets	2 weeks of May 2013	MERS-Cov	Tweets Monitoring Platforms with ML Approach to Automatically Detect and Classify Healthcare Tweets Associated with MERS-COV	ML Based Models
Singh et al. (2018)	N/A	12 Study	2010–2017	Epidemic	Discussing Sentiment Analysis ML based with Pandemics	ML Based Models
Jain and Kumar (2018)	Twitter	10,000 Tweets	September 2016 To November 2016	Outbreak	Tracking of Real-Time Public Sentiments for Outbreak	Hybrid Based
Ji et al. (2015)	Twitter	Around 15 Million Tweets.	From 13 March 2014 to 29 June 2014	Infectious Diseases	Measuring Public Concern Using a Two-Step Sentiment Classification Approach	Hybrid Based Models
Pollacci et al. (2017)	Twitter	3,718 Publicly Available Tweets	14–20 August 2015	Epidemic	Sentiment Spreading Based Approach For Epidemic	Hybrid Based Models
Choi et al. (2017)	News Media Twitter	86 Million Words	70 Days	MERS-Cov	Monitoring and Understanding Responses Public Outbreak of Infectious Diseases	Hybrid Based Models
Ji et al. (2016)	Twitter	10,657 sentences	June 2001 to May 2002	Epidemic	Epidemic Sentiment Monitoring System	Hybrid Based Models
Ji, Chun, and Geller (2013)	Twitter	126,833 tweets	26–28 September 2011 9–18 October 2011	Epidemic	Epidemic Sentiment Monitoring System	Hybrid Based Models
S. Li et al. (2020)	Weibo	posts from 17,865 users	13–26 January 2020	COVID-19	Exploring the Impacts of COVID-19 on People's Mental Health to Assist Policy and Provide Timely Services to Infected Populations	Hybrid Based Models
Jain and Kumar (2015)	Twitter	91,495 tweets	1 February 2015 to 31 March 2015	N1H1	Method for Tracking the N1H1 Using the Contents of Twitter	Hybrid Based Models
L. Li et al. (2020)	Weibo	367,462 posts	30 December 2019 to 1 February 2020	COVID-19	Characterising the Propagation of Situational Information in Social Media During COVID-19 Epidemic:	Hybrid Based Models
Chaudhary and Naaz (2017)	Facebook Twitter	N/A	1 April 2017 to 30 June 2017	Epidemics	Discussion of the Importance of Utilising Big Data Means Like Sentiment Analysis	Individuals
Gayo-Avello et al. (2013)	Social Media	N/A	N/A	Outbreaks	Discussion for Social Media SM Predictive Power	Individuals
Seltzer, Horst-Martz, Lu, and Merchant (2017)	Instagram	141,161 posts	May–August 2016	Zika Virus	Analysing Image Sentiment from Instagram for the Spread of Zika Virus and Information Dissemination	Individuals
Al-garadi, Khan, Varathan, Mujtaba, and Al-Kabsi (2016)	Digital Databases	20 Study	2005–2014	Pandemics	A Systematic Literature Search for Studies with the Primary Aim of Using OSN to Detect and Track a Pandemic	Individuals

and understand the emotional response of the public to a widespread outbreak of infectious diseases. The methodology is based on the analysis of the massive media outlet data collected in the 2015 nationwide outbreak of MERS-CoV in Korea. Data were collected, processed and analysed with more than 86 million words for composing a methodology of ML and information theoretic approaches. The techniques for extracting emotions from emoticons and Internet slang were also included. The approach provided an efficient method for the rapid monitoring of public reaction to an infectious disease on a national scale and revealed useful information for the timely control of the disease. In another work by Ji et al. (2016), a two-step sentiment classification method was developed by combining clue-based labelling and ML methods by automatically labelling the training datasets and then building classifiers for personal tweets and classifiers for tweet sentiments. In addition, a computational intelligence approach was developed for an epidemic sentiment monitoring system (ESMOS) to automatically analyse the linguistic expressions that convey the subjective expressions and sentiment polarity of emotions, feelings, opinions and personal attitudes. The method could be generalised to other topical domains, such as mental health monitoring and crisis management. In Ji, Chun, and Geller (2013), an ESMOS that provides tools for visualising users' tweets and posts regarding the concerns towards different diseases was discussed; the system could help public health officials identify the progression and peaks of concerns for a disease in space and time, which enables the implementation of appropriate preventive actions to mitigate these diseases. A classification approach was utilised to identify the negative sentiment of personal health tweets and measure the degree of concern (DOC) for the daily monitoring of public sentiments towards a disease. In this work, different ML methods were applied to classify sentiments of Twitter users regarding diseases into personal and neutral tweets and differentiate negative tweets from neutral personal tweets. In the experiments, multinomial NB generally achieved the best results and required significantly less time to build the classifier than other methods. Another interesting work within this category by S. Li et al. (2020) explored the mental impact of COVID-19 on the Chinese via predictive modelling (i.e., ML) and sentiment analysis. In this regard, a text mining system was developed from the Chinese Academy of Sciences to extract content features, followed by the Chinese word segmentation tool and psychoanalytic dictionary towards categorising microblog content into linguistic annotations. The latter included various emotions, such as positive, negative and angry emotions, which could reflect the actual experiences of people. People showed other negative emotions (anxiety, depression and indignation) and less positive emotions after the declaration of COVID-19. Another study by Jain and Kumar (2015) proposed a sentiment analysis method to track N1H1 in India by using Twitter content. Sentiment analysis and count-based techniques were applied to data sets that contain 91,495 tweets and queries to examine important issues relevant to the presence of N1H1 symptoms. Data were preprocessed, and classification was then carried out for relevant word-related searches. Each word was considered a feature, and various classification techniques, such as the SVM, NB, random forest (RF) and decision tree, were applied. The findings of the classification confirmed that the SVM classifier produced better results than those by the other classifiers with an F-measure of 0.72. Another study by L. Li et al. (2020) discussed the use of lexical- and ML-based approaches in the fight against COVID-19 by utilising natural language processing (i.e., sentiment analysis) to classify social media content into several types and choose Weibo as a study case. A total of 367,462 posts were used to extract relevant features that were applied to train ML algorithms, such as SVM, NB and RF, to learn the types of unlabelled data based on labelled data. COVID-19-related information was classified into seven types of situational information, including emotional, perception and affiliation factors. The best results were achieved by using the best-performing RF classifier. Previous studies discussed lexicon-based literature with ML-based literature as a hybrid approach from the taxonomy discussion. Fig. 4 also illustrates the usage of this

approach in relation to COVID-19, $n = 2$; epidemic, $n = 3$; MERS-CoV, $n = 1$; outbreak, $n = 1$; infectious disease, $n = 1$ and N1H1, $n = 1$.

3.4. Individuals

The last category addresses individual topics with sentiment analysis. The sections contain a total of $n = 4/28$ articles. The first work by Chaudhary and Naaz (2017) discussed the importance of utilising big data means, such as sentiment analysis, from social media to address new epidemics and described that slow rate responses to these occurrences might prevent countries from contracting these diseases. The study also raised the importance of including unstructured data sources, such as social media (i.e., Facebook or Twitter), by governments and performing opinion analysis while publishing national health and epidemic outbreak advisories in the near future. Furthermore, the study argued that traditional case-based reporting may lack efficiency and promptness in disease outbreak reporting and may easily be achieved by harnessing the hidden potential of health data generated online in large volumes with greater dynamics. Another work by Gayo-Avello et al. (2013) described the power of social media and its consolidated knowledge and subsequently evaluated predictions in various areas, such as disease outbreaks, product sales, stock market volatility and election outcome predictions. In this work, the literature was systematically reviewed to identify relevant empirical studies, and the selected studies were analysed and synthesised in the form of a proposed conceptual framework, which was thereafter applied to further examine the literature and provide new insights in the field. The review concluded that most studies have supported the predictive power of social media but more than one-third of these studies have inferred the predictive power without employing predictive analytics. In addition, the analysis suggested the distinct need for other advanced sentiment analysis methods and approaches to identify search terms for the collection and filtering of raw social media data. In another work by Seltzer, Horst-Martz, Lu, and Merchant (2017), sentiment analysis was conducted from different perspectives, which include images rather than text. More than 141,161 Instagram posts were considered to explore how this image platform was applied to spread information related to the spread of Zika virus. The images in the related posts for the sentiment analysis were categorised into four codes, namely, humour, fear, positive, negative and neutral. The experimental result showed that the majority of the image sentiments were split amongst humour (26%), fear (29%), positive (21%) and negative (22%). The conclusion was that the fear of Zika virus transmission was dominant in the majority of negative posts and that utilising social media platforms with image content could be employed as a tool to gauge public sentiment during a public health emergency. In the last work by Al-garadi, Khan, Varathan, Mujtaba, and Al-Kabsi (2016), a systematic review aimed to examine how social media platforms were utilised to track pandemics. Studies between 2005 and 2014 were reviewed, and the findings indicated that the popularity and proliferation of online social networks (OSNs) could contribute to the development of excellent real-time pandemic surveillance. However, this work was associated with excluded articles, which did not employ data from the social network as the data source for opinion mining and sentiment analysis purposes and actually reflect the main essence of this paper. The work focused on the role of OSN in surveillance but disregarded people's emotions and sentiments. Previous studies excluded sentiments with any of the three major categories. Moreover, the remaining papers addressed sentiment analysis with the topic and usage in relation to outbreaks, $n = 1$; epidemic, $n = 1$; Zika virus, $n = 1$; and pandemic, $n = 1$.

4. Application languages

This section illustrates and explains the languages that are utilised in the sentiment analysis studies mentioned in this review. All publications included in this manuscript were written in English as part of the

systematic review criteria. Nevertheless, all the applications of sentiment analysis were conducted in English and other languages. Therefore, this section aims to present them as shown in Table 4.

In Table 4, English was the dominant language utilised for the sentiment analysis research, with $n = 17/28$ studies analysing English data, followed by Arabic, with a total of $n = 4/28$ studies that analyse Arabic data with sentiment analysis, Chinese with $n = 3/28$, India with two studies, and Korean and Filipino with one study for each language. However, languages such as English as the dominant language do not always mean that the study was conducted in a Western country, such as the USA. Some of these studies were conducted in non-English speaking countries, but English tweets were utilised to analyse the population's sentiment, which is clearly indicated by the country of the first author. Another potential reason is the place of origin of the infectious disease. An example is MERS-CoV, which was first detected in Saudi Arabia in 2012. Thus, we obtained research in Arabic, as this case started in this region and attracted many researchers to work with the language of this domain. However, COVID-19 has been extremely widespread compared with dreaded epidemics and pandemics, such as 'The Great Influenza' (Spanish flu of 1918) or the Black Death (form of bubonic plague), which spread worldwide (Bhat et al., 2020). During these times, no computerised technologies or analytical methods, such as sentiment analysis, were available to study these phenomena. Thus far, further research is excessively available in many languages, and different applications will be conducted using sentiment analysis because of the global impact of COVID-19.

5. Discussions

This section aims to highlight and discuss two major components of this study. After the protocol and data extraction, a pattern of extracted literature information was apparent. Most of the related literature was deemed suitable with one of the two major discussion components: (1) challenges in which issues and problems faced by previous and existing academics and researchers are reported and (2) motivations in which the significance and benefits of the topic are reported and ideas can explain why these topics were pursued by researchers.

Table 4
Application languages of sentiment analysis.

Language Application	Author	Number of Publications	References
English	USA	7	Lim et al. (2017); Chung et al. (2015); Ji et al. (2015, 2016, 2013); Culotta (2010); Seltzer et al. (2017)
	India	3	Chaudhary and Naaz (2017); Bhat et al. (2020); Jain and Kumar (2015)
	Australia	1	K. Ali et al. (2017)
	Republic of Korea	1	E. H.-J. Kim et al. (2016)
	Italy	1	Pollacci et al. (2017)
	Greece	1	Gayo-Avello et al. (2013)
	Singapore	2	Lwin et al. (2020); Raamkumar et al. (2020)
	Malaysia	1	Al-garadi et al. (2016)
	Qatar	1	A. Ali et al. (2013)
	Jordan	1	Baker et al. (2020)
Arabic	Saudi Arabia	2	Almazidy et al. (2016); Zarrad et al. (2014)
	China	3	L. Li et al. (2020); Deng et al. (2015); S. Li et al. (2020)
Chinese	China	3	Choi et al. (2017)
Korean	Korea	1	Jain and Kumar (2018); Singh et al. (2018)
Indian	India	2	Pastor (2020)
Filipino	Philippines	1	

5.1. Challenges

Issues and challenges are amongst the most common academic dilemmas. Whether they are directly related to a researcher's area of interest or have an indirect impact, further studies are needed to address these dilemmas and advance science in the designated domain. In sentiment analysis associated with infectious diseases, challenges are distributed in different categories and grouped based on some common characteristics to facilitate understanding for future research.

5.1.1. Challenges related to data

Data are the most important aspect in any study and the most significant effector that controls the analysis, findings and all aspects of research elements. In sentiment analysis, data have been repeatedly shown to be an active research area (Chaudhary & Naaz, 2017) that can offer valuable insights into new diseases (Chaudhary & Naaz, 2017) and outbreaks (Gayo-Avello et al., 2013) with the understanding of natural and varied settings. Moreover, data are fast-growing topics in textual detection (*i.e.*, *sentiment*) from a variety of broadcast, news press and social media sites (Al-garadi et al., 2016; E. H.-J. Kim et al., 2016) that are utilised to analyse opinions (Zarrad et al., 2014). All the challenges in this regard are linked to either data processing or collection. In the former, the identified literature highlights issues such as the noisy nature of data from social media sites (Al-garadi et al., 2016; K. Ali et al., 2017; Almazidy et al., 2016; Chaudhary & Naaz, 2017; Gayo-Avello et al., 2013) or insufficiency (Pollacci et al., 2017; Seltzer et al., 2017; Zarrad et al., 2014). Other authors highlighted data processing issues with regard to the irrelevancy of data (K. Ali et al., 2017) or the premise that data have different types and formatting styles (Al-garadi et al., 2016; K. Ali et al., 2017; Chaudhary & Naaz, 2017; Jain & Kumar, 2018, 2015). Aside from other scholars who agreed that processing data from social media requires excellent processing skills (Gayo-Avello et al., 2013), the need for a high volume of analysis procedures is acknowledged (Al-garadi et al., 2016; Chaudhary & Naaz, 2017; Jain & Kumar, 2018; Zarrad et al., 2014). Conversely, collecting data on a user's opinions regarding a specific subject is an immense burden and challenging process (Al-garadi et al., 2016; Zarrad et al., 2014) as social media platforms proliferate, and thus, quickly multiply the amount of data (Jain & Kumar, 2018). This problem is also attributed to the difficulty in defining keywords to identify the desired data (Chaudhary & Naaz, 2017).

5.1.2. Challenges related to social media platforms

As a growing online platform for sharing opinions and ideas, social media provides many opportunities for decision makers to understand public emotions (S. Li et al., 2020; Pastor, 2020; Raamkumar et al., 2020; Seltzer et al., 2017). Social media enables users to promptly express their emotions (Chung et al., 2015). Many researchers around the globe have explored the usage of social media in diverse domains (K. Ali et al., 2017), and among their most important domain is their crucial role during disease outbreaks (Almazidy et al., 2016; Chaudhary & Naaz, 2017). Social media analysis is a promising area as generated data are highly dynamic and useful for real-time trends (A. Ali et al., 2013; Chaudhary & Naaz, 2017). This approach can provide a rapid and effective monitoring mechanism for public health on a large scale at a low cost (Al-garadi et al., 2016; Choi et al., 2017; Culotta, 2010; Jain & Kumar, 2015; Pastor, 2020). Among the most well-known social media websites that are highly utilised in sentiment analysis are Facebook, Twitter, Reddit, Instagram, and news forums (K. Ali et al., 2017). Even these valuable assets have various challenges that can hinder their ultimate application. Researchers in the sentiment analysis domain identified issues and challenges in the social media platform with regard to reliability (Jain & Kumar, 2018) and authenticity (Almazidy et al., 2016; Chaudhary & Naaz, 2017; Zarrad et al., 2014). Other researchers discussed the challenges in social media platforms because of the cadence of content capability limitation (E. H.-J. Kim et al., 2016; Pollacci et al.,

2017), possible exaggeration (Choi et al., 2017) and difficulty in understanding sentiments, especially in cases such as disease outbreaks (Chung et al., 2015; Ji et al., 2013) or different sources (K. Ali et al., 2017; Pollacci et al., 2017).

5.1.3. Challenges related to the community

The community in this section is dedicated to any involvement of individuals or parties from various fields or agencies. In scanning the challenges associated with sentiment analysis and infectious diseases and outbreaks, in addition to viruses, epidemics and pandemics, two parties were mostly directed: (1) decision makers and (2) the scientific community. Decision makers' challenges include identifying people's sentiments on a subject and their beliefs (Zarrad et al., 2014) towards public health policy decision-making (Chung et al., 2015; S. Li et al., 2020; Lwin et al., 2020) in extremely serious cases, such as disease outbreaks (Chung et al., 2015). In addition, this dilemma will challenge decision makers to take serious measures without delay (Ji et al., 2015) and control the spread of an event or even quarantine confirmed cases of infectious diseases (Bhat et al., 2020; Choi et al., 2017; Ji et al., 2013; Singh et al., 2018). Aside from decision makers, the scientific community has an equally important role in fighting these outbreaks via sentiment analysis; however, fully utilising the abilities of the community remains elusive because of a lack of scientific reference work (Chung et al., 2015), the availability of studies in certain languages only (Baker et al., 2020) or the absence of relevant research with the latest trends (Ji et al., 2015).

5.2. Motivation

Researchers and academics are drawn to their respected fields for various scientific reasons. Some researchers are encouraged by the significance of topics, while some researchers are enthusiasts who are attracted by the promised benefits of the topics. For sentiment analysis associated with infectious diseases, different levels of significance and motivation show researchers' solicitude in this area in terms of two classes, namely, disease mitigation and data analysis. These classes are summarised and grouped in the following subsections.

5.2.1. Motivation related to disease mitigation

All countries periodically face new diseases (*i.e.*, epidemics), but the slow rate at which countries can respond and contain these diseases shatters and victimises the country's population on massive scales (Chaudhary & Naaz, 2017), which produces massive effects on emotions (Choi et al., 2017) and creates widespread panic (Chung et al., 2015). Using traditional approaches to mitigate these kinds of diseases fails to address the challenge (Choi et al., 2017); nevertheless, the use of intelligent and fast channels is successful, and social media has a crucial role as a primary channel of communication during disease outbreaks (Almazidy et al., 2016; Pastor, 2020). In carefully analysing research efforts that are devoted to academic literature on the sentiment analysis and opinion mining domain in mitigating diseases, outbreaks and infectious diseases, sentiment analysis has shown its significance via its four main important aspects, namely, (1) *monitoring*, (2) *discovery*, (3) *news sharing* and (4) *policies*. For monitoring, related motivations focused on tracking human activities based on geographical locations (Al-garadi et al., 2016; K. Ali et al., 2017; Ji et al., 2013), containing the spread of contagious diseases (Culotta, 2010; Deng et al., 2015; Ji et al., 2013, 2016, 2015; Singh et al., 2018) and avoiding public concerns and panic (Ji et al., 2016; S. Li et al., 2020; Lwin et al., 2020; Seltzer et al., 2017). For discovery, the literature discussed two main aspects: fast disease discovery (Al-garadi et al., 2016; Baker et al., 2020; Culotta, 2010) and early disease discovery (A. Ali et al., 2013; Jain & Kumar, 2018). For news sharing, predictability is acknowledged due to the capability of sentiment analysis by sharing news about diseases and outbreaks via social media (Almazidy et al., 2016); similarly, other emergencies (K. Ali et al., 2017; S. Li et al., 2020) can be easily shared via social media. For

policies, governments can largely subject social media to sentiment analysis to control an epidemic (Jain & Kumar, 2015; Pastor, 2020; Raamkumar et al., 2020) and gain adequate time to act (Chaudhary & Naaz, 2017; S. Li et al., 2020) towards handling epidemic outbreaks by promptly issuing health advisories (Chaudhary & Naaz, 2017; S. Li et al., 2020), such as social distancing and satisfactory hygiene habits (Raamkumar et al., 2020), or even making decisions to handle emergency accident cases (Deng et al., 2015; L. Li et al., 2020; Seltzer et al., 2017).

5.2.2. Motivation related to data analysis

Data analysis has various important roles in all research areas, including infectious disease and prevention (Chaudhary & Naaz, 2017; Gayo-Avello et al., 2013; S. Li et al., 2020). The role of data analysis may not be as risky as the role of physicians and medical doctors, who face an overwhelming number of cases daily and deal with the latter at the expense of their health. Nevertheless, analysing the data for the sentiment of infectious diseases is an active research trend in natural language processing and data mining (Zarrad et al., 2014). This analysis can have a useful role in medical decisions, such as those relevant in disease spread (Pastor, 2020). Additionally, this type of analysis carries numerous motivations, as discussed in the literature. The data analysis of sentiments can be beneficial for various reasons, including the simple, fast and effective means for studying public data sentiment on diseases (Culotta, 2010; L. Li et al., 2020; Pollacci et al., 2017) and infection spread (Jain & Kumar, 2018); efficiently preventing an epidemic (Ji et al., 2015; Raamkumar et al., 2020) and outbreak (Ji et al., 2013) by harnessing the hidden potential of health-related data generated online in large volumes (Chaudhary & Naaz, 2017; Culotta, 2010; Singh et al., 2018); visualising infectious disease transmission (Chung et al., 2015); keeping track of trends concerning public health (Jain & Kumar, 2018; Ji et al., 2015; Seltzer et al., 2017), which can contribute to diagnosis (Baker et al., 2020), medical treatments (Ji et al., 2013; Lim et al., 2017), vaccination campaigns and strategies (Chaudhary & Naaz, 2017; L. Li et al., 2020), decision-making processes (Jain & Kumar, 2018; Ji et al., 2013; Zarrad et al., 2014) or even the development of a regularly updated atlas of infectious diseases (Chaudhary & Naaz, 2017; S. Li et al., 2020); and assisting in urgent interventional care (Bhat et al., 2020; Pastor, 2020; Seltzer et al., 2017) and proper management of time and resources (Jain & Kumar, 2018).

6. Contributions

This section aims to describe the contributions of this review in terms of two different aspects: theoretical contributions and practical contributions.

6.1. Theoretical contribution

This study introduces several theoretical contributions. First, the discussion in this study serves as a guide and important source of knowledge related to sentiment analysis applications in infectious diseases. A protocol of a systematic literature review was adopted to summarise the included studies within the designated criteria and reintroduce them in a new theme, which helps researchers examine these studies from different perspectives. Researchers can ponder the discussion section and consider the challenges that their peers previously faced. They can gain a glimpse of problems that require attention and direct their efforts towards addressing these issues in the future. Another aspect is related to motivations that contribute to the theoretical knowledge introduced in this study by showing the significance of conducting these studies, not only for solving problems but also for providing potential benefits and advantages, such as mitigating disease-reducing costs. Another theoretical contribution is linked to demographic statistics that introduce the recent statistics of the studies included in this review. These statistics can help researchers obtain a

clear idea of recent publications across the years and various sources as future references.

6.2. Practical contribution

All researchers across academic disciplines explore topics that are similar or related to infectious diseases because of *COVID-19*. This study practically contributes to current data by presenting information in Table 3, which aids in the design of methodological aspects based on previous research findings. This task helps determine highly important information, such as the best social media platform that can be employed for sentiment analysis and data mining and the method of data extraction. In addition, this study can help determine data volume by presenting other methodological aspects linked to this part and providing a basis for planning future studies. All the aspects discussed in Table 3 can help researchers address their practical concerns related to sentiment analysis. Another contribution of this study is the utilisation of SLR to improve the understanding of the topic. This review illustrates how each study is selected and analysed, *i.e.*, from searching to choosing keywords and providing a discussion.

7. Authors' claims and discussion

COVID-19 as an infectious disease that remains ambiguous; an accurate prediction can only be obtained when the pandemic ends (Hamzah et al., 2020). The pandemic is substantially influenced by each country's policy and social responsibility (Anderson, Heesterbeek, Klinkenberg, & Hollingsworth, 2020). Data transparency is crucial inside a government (Alwan et al., 2020), and avoiding the dissemination of unverified news and staying calm in this situation are our responsibility (Yusof, Muuti, Ariffin, & Tan, 2020). Using sentiment analysis to fight this pandemic illustrates the importance of information dissemination, which can help improve response times and establish advanced planning to reduce risks that can be influenced by social media. Information spread via social media has an important role in empowering citizens during a pandemic and witnessing public reaction about fake news during this time. These parameters can be measured using data from social media. Different languages, people and emotions call for an end to this pandemic. All sentiments, expressions and opinions can contribute to an improvement in the findings related to a pandemic similar to *COVID-19* and providing timely information to the public. In addition, these sentiment analyses on social media platforms assist governments and authorities in disseminating verified articles, providing updates, advocating effective personal hygiene and promoting social responsibility in spreading awareness to the public by providing scientific-based data analysis, prediction and verified news. Studies have scarcely examined the presence of pandemics such as this pandemic via sentiment analysis (Al-garadi et al., 2016; Jain & Kumar, 2015; Raamkumar et al., 2020) and were identified only in terms of the number of related studies included in this review, which is equivalent to 10 years of research efforts. This phenomenon might be attributed to the premise that the scale of previous epidemics was considerably smaller with a lower number of infections and rate of spread. This pandemic has highlighted the extent of our vulnerability and revealed that we have the means and technologies, especially technological tools such as sentiment analysis, for its mitigation and prevention. We can make several claims by considering the following current situations:

- A massive increase in the number of studies on *COVID-19* topics is expected across all scientific disciplines, and sentiment analysis is expected to increase with different languages and applications.
- Social media is a highly important news medium, and this power can create panic, false news and harmful practices on a large scale if the use of social media is not properly controlled.

8. Conclusion

In this research, studies about sentiment analysis in the presence of infectious diseases, outbreaks, epidemics and pandemics over a 10-year period (1 January 2010 to 30 June 2020) were systematically reviewed. The research motivation of this work was the massive spread of *COVID-19*. *COVID-19* as an infectious disease remains vague as its literature and cases are proliferating massively; consequently, reporting updated information is nearly impossible. Furthermore, accurate information can only be obtained when the pandemic ends. Further studies should focus on the role of social media and sentiment analysis when a similar incident recurs. This systematic review addressed the main highlights: the protocol that explains how the last set of articles was chosen, a taxonomy analysis of current papers in the field and previous research efforts in the form of challenges and motivations. Despite the relatively low number of studies in this field, current data are essential for fighting similar outbreaks in the future and to stand in the face of these crises, not only as medical doctors and researchers but also as scientists from all domains, communities and decision-making bodies. In addition, studies should consider how our respected area can be an asset in the near future. From the perspectives of computer science, integrating other technologies, such as AI, ML and different analysis procedures, can contribute to making a difference. We all stand as one entity in the hope that technologies and science will help prevent a re-emergence of diseases such as *COVID-19*.

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. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.eswa.2020.114155>.

References

- Ain, N., Vaia, G., DeLone, W. H., & Waheed, M. (2019). Two decades of research on business intelligence system adoption, utilization and success—A systematic literature review. *Decision Support Systems*, 125, 113113.
- Ainin, S., Feizollah, A., Anuar, N. B., & Abdullah, N. A. (2020). Sentiment analyses of multilingual tweets on halal tourism. *Tourism Management Perspectives*, 34, 100658.
- Al-Anzi, F. S., & AbuZeina, D. (2017). Toward an enhanced Arabic text classification using cosine similarity and Latent Semantic Indexing. *Journal of King Saud University—Journal of Computer and Information Sciences*, 29, 189–195.
- Al-garadi, M. A., Khan, M. S., Varathan, K. D., Mujtaba, G., & Al-Kabsi, A. M. (2016). Using online social networks to track a pandemic: A systematic review. *Journal of Biomedical Informatics*, 62, 1–11.
- Alamoody, A., Garfan, S., Zaidan, B., Zaidan, A., Shuwandy, M. L., Alaa, M., ... Albahri, O. (2020). A systematic review into the assessment of medical apps: Motivations, challenges, recommendations and methodological aspect. *Health and Technology*, 1–17.
- Alamoody, A., Zaidan, B., Zaidan, A. A., Samuri, S. M., Ismail, A. R., Zughoul, O., ... Chyad, M. (2019). A review of data analysis for early-childhood period: Taxonomy, motivations, challenges, recommendation, and methodological aspects. *IEEE Access*, 7, 51069–51103.
- Ali, A., Magdy, W., & Vogel, S. (2013). A tool for monitoring and analyzing healthcare tweets. In *Proceedings of the ACM SIGIR workshop on health search & discovery* (p. 23).
- Ali, K., Dong, H., Bouguettaya, A., Erradi, A., & Hadjidi, R. (2017). In *Sentiment analysis as a service: a social media based sentiment analysis framework* (pp. 660–667). IEEE.
- Aljuaid, H., Ifitkhar, R., Ahmad, S., Asif, M., & Afzal, M. T. (2020). Important citation identification using sentiment analysis of In-text citations. *Telematics and Informatics*, 101492.
- Almazidy, A., Althani, H., & Mohammed, M. J. P. C. S. (2016). Towards a disease outbreak notification framework using Twitter mining for smart home dashboards. *Procedia Computer Science*, 82, 132–134.
- Alwan, N. A., Bhopal, R., Burgess, R. A., Colburn, T., Cuevas, L. E., Smith, G. D., ... Gilthorpe, M. S. (2020). Evidence informing the UK's *COVID-19* public health response must be transparent. *Lancet (London, England)*, 395, 1036–1037.
- Anderson, R. M., Heesterbeek, H., Klinkenberg, D., & Hollingsworth, T. D. (2020). How will country-based mitigation measures influence the course of the *COVID-19* epidemic? *The Lancet*, 395, 931–934.

- Appel, G., Grewal, L., Hadi, R., & Stephen, A. T. (2020). The future of social media in marketing. *Journal of the Academy of Marketing Science*, 48, 79–95.
- Baker, Q. B., Shatnawi, F., Rawashdeh, S., Al-Smadi, M., & Jararweh, Y. (2020). Detecting epidemic diseases using sentiment analysis of Arabic Tweets. *Journal of Universal Computer Science*, 26, 50–70.
- Bhat, M., Qadri, M., Noor-ul-Asrar Beg, M. K., Ahanger, N., & Agarwal, B. J. B. (2020). Behavior, & immunity, sentiment analysis of social media response on the Covid19 outbreak. *Brain, Behavior, and Immunity*.
- Burgers, C., Brugman, B. C., & Boeynaems, A. (2019). Systematic literature reviews: Four applications for interdisciplinary research. *Journal of Pragmatics*, 145, 102–109.
- Chaudhary, S., & Naaz, S. (2017). In *Use of big data in computational epidemiology for public health surveillance* (pp. 150–155). IEEE.
- Choi, S., Lee, J., Kang, M.-G., Min, H., Chang, Y.-S., & Yoon, S. J. M. (2017). Large-scale machine learning of media outlets for understanding public reactions to nation-wide viral infection outbreaks. *Methods*, 129, 50–59.
- Chung, W., He, S., & Zeng, D. (2015). emood: Modeling emotion for social media analytics on Ebola disease outbreak. *International Conference on Information Systems*.
- Crawford, K. J. C. (2009). Following you: Disciplines of listening in social media. *Continuum*, 23, 525–535.
- Culotta, A. (2010). Towards detecting influenza epidemics by analyzing Twitter messages. In *Proceedings of the first workshop on social media analytics* (pp. 115–122).
- DeNardis, L., & Hackl, A. M. J. T. P. (2015). Internet governance by social media platforms. *Telecommunications Policy*, 39, 761–770.
- Deng, X. L., Tang, Y. Q., & Huang, Y. H. (2015). Opinion mining for emergency case risk analysis in spark based distributed system. In *Proceedings of the 1st ACM SIGSPATIAL international workshop on the Use of GIS in emergency management* (pp. 1–8).
- Diedrichs, C. P. J. L. C., Acquisitions, & Services, T. (2000). Highlights of the second European ScienceDirect user conference. 24, 391–402.
- Elaish, M. M., Shuib, L., Ghani, N. A., Yadegaridehkordi, E., & Alaa, M. J. I. A. (2017). Mobile learning for English language acquisition: Taxonomy, challenges, and recommendations. *IEEE Access*, 5, 19033–19047.
- Elder, E., Johnston, A. N., Wallis, M., & Crilly, J. J. I. E. N. (2020). The demoralisation of nurses and medical doctors working in the emergency department: A qualitative descriptive study. *International Emergency Nursing*, 100841.
- Epidemiol, C. P. M. A. (2020). An update on the epidemiological characteristics of novel coronavirus pneumonia (COVID-19). *Preventive Medicine Association*, 41, 139–144.
- Eysenbach, G. (2003). SARS and population health technology. *Journal of Medical Internet Research*, 5, e14.
- Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PubMed, Scopus, web of science, and Google scholar: Strengths and weaknesses. *The FASEB Journal*, 22, 338–342.
- Gayo-Avello, D., Metaxas, P. T., Mustafaraj, E., Strohmaier, M., Schoen, H., Gloor, P., ... Tarabanis, K. J. I. R. (2013). Understanding the predictive power of social media. *Internet Research*.
- Gies, T. J. L. P. (2018). The ScienceDirect accessibility journey: A case study. *Learned Publishing*, 31.
- Goldschmidt, K. (2020). The COVID-19 pandemic: Technology use to support the wellbeing of children. *Journal of Pediatric Nursing*.
- Griffin, L. (2002). IEEE Xplore. Version 1.3.
- Hamzah, F. B., Lau, C., Nazri, H., Ligot, D., Lee, G., & Tan, C. (2020). CoronaTracker: Worldwide COVID-19 outbreak data analysis and prediction. *Bull World Health Organ*, 1, 32.
- Hassan, S.-U., Aljohani, N. R., Idrees, N., Sarwar, R., Nawaz, R., Martínez-Cámara, E., ... Herrera, F. J. K.-B. S. (2020). Predicting literature's early impact with sentiment analysis in Twitter. *Knowledge-Based Systems*, 192, 105383.
- Jain, V. K., & Kumar, S. (2018). Effective surveillance and predictive mapping of mosquito-borne diseases using social media. *Journal of Computational Science*, 25, 406–415.
- Jain, V. K., & Kumar, S. (2015). An effective approach to track levels of influenza-A (H1N1) pandemic in India using twitter. *Procedia Computer Science*, 70, 801–807.
- Jansen, B. J., Sobel, K., & Cook, G. (2010). Gen X and Ys attitudes on using social media platforms for opinion sharing. In *CHI'10 extended abstracts on human factors in computing systems* (pp. 3853–3858).
- Ji, X., Chun, S. A., & Geller, J. (2013). In *Monitoring public health concerns using twitter sentiment classifications* (pp. 335–344). IEEE.
- Ji, X., Chun, S. A., & Geller, J. (2016). Knowledge-based tweet classification for disease sentiment monitoring. In *Sentiment analysis and ontology engineering* (pp. 425–454). Springer.
- Ji, X., Chun, S. A., Wei, Z., Geller, J., & Mining. (2015). Twitter sentiment classification for measuring public health concerns. *Social Network Analysis and Mining*, 5, 13.
- Kim, E. H.-J., Jeong, Y. K., Kim, Y., Kang, K. Y., & Song, M. (2016). Topic-based content and sentiment analysis of Ebola virus on Twitter and in the news. *Journal of Information Science*, 42, 763–781.
- Kim, K.-S., Sin, S.-C. J., & Yoo-Lee, E. Y. (2014). Undergraduates' use of social media as information sources. *WKWSCI Journal Articles*.
- Kushwah, S., Dhir, A., Sagar, M., & Gupta, B. (2019). Determinants of organic food consumption. A systematic literature review on motives and barriers. *Appetite*, 104402.
- Li, L., Zhang, Q., Wang, X., Zhang, J., Wang, T., & Gao, T.-L. (2020). Characterizing the propagation of situational information in social media during covid-19 epidemic: A case study on weibo. *IEEE Transactions on Computational Social Systems*, 7, 556–562.
- Li, S., Wang, Y., Xue, J., Zhao, N., Zhu, T. J., & health, p. (2020). The impact of COVID-19 epidemic declaration on psychological consequences: a study on active Weibo users. *International Journal of Environmental Research and Public Health*, 17, 2032.
- Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., & Ioannidis, J. P.o.c.e. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *Journal of Clinical Epidemiology*, 62, e1–e34.
- Lim, S., Tucker, C. S., & Kumara, S. (2017). An unsupervised machine learning model for discovering latent infectious diseases using social media data. *Journal of Biomedical Informatics*, 66, 82–94.
- Liu, R., Han, H., Liu, F., Lv, Z., Wu, K., Liu, Y., ... Zhu, C. J. C. C. A. (2020). Positive rate of RT-PCR detection of SARS-CoV-2 infection in 4880 cases from one hospital in Wuhan, China, from Jan to Feb 2020. *Clinica Chimica Acta*.
- Loureiro, S. M. C., Romero, J., & Bilro, R. G. (2019). Stakeholder engagement in co-creation processes for innovation: A systematic literature review and case stud. *Journal of Business Research*.
- Lwin, M. O., Lu, J., Sheldenkar, A., Schulz, P. J., Shin, W., Gupta, R., Yang, Y., & surveillance. (2020). Global sentiments surrounding the COVID-19 pandemic on Twitter: analysis of Twitter trends. *JMIR Public Health and Surveillance*, 6, e19447.
- McMullan, L. K. (2020). Clinical trials in an Ebola outbreak seek to find an evidence-based cure. *EBioMedicine*.
- Meho, L. I., & Yang, K. (2007). technology. Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology*, 58, 2105–2125.
- Pan, X., Ojcius, D. M., Gao, T., Li, Z., Pan, C., & Pan, C. J. M. (2020). Infection. Lessons learned from the 2019-nCoV epidemic on prevention of future infectious diseases. *Microbes and Infection*, 22, 86–91.
- Pastor, C. K. (2020). Sentiment analysis of filipinos and effects of extreme community quarantine due to Coronavirus (Covid-19) Pandemic. *Journal of Critical Reviews*.
- Pollacci, L., Sirbu, A., Giannotti, F., Pedreschi, D., Luchese, C., & Muntean, C. I. (2017). In *Sentiment spreading: An epidemic model for lexicon-based sentiment analysis on twitter* (pp. 114–127). Springer.
- Raamkumar, A. S., Tan, S. G., & Wee, H. L. (2020). Measuring the outreach efforts of public health authorities and the public response on facebook during the COVID-19 pandemic in early 2020: Cross-country comparison. *Journal of Medical Internet Research*, 22, e19334.
- Reyes-Menendez, A., Saura, J. R., & Filipe, F. J. H. (2020). Marketing challenges in the# MeToo era: Gaining business insights using an exploratory sentiment analysis. *Heliyon*, 6, e03626.
- Rodriguez, R. G., das Dores, R. M., Camilo-Junior, C. G., & Rosa, T. C. (2016). SentiHealth-Cancer: A sentiment analysis tool to help detecting mood of patients in online social networks. *International Journal of Medical Informatics*, 85, 80–95.
- Rotaru, A. S., & Vigiocco, G. (2020). Constructing semantic models from words, images, and emojis. *Cognitive Science*, 44, e12830.
- Seltzer, E., Horst-Martz, E., Lu, M., & Merchant, R. (2017). Public sentiment and discourse about Zika virus on Instagram. *Public Health*, 150, 170–175.
- Singh, R., Singh, R., & Bhatia, A. (2018). Sentiment analysis using Machine Learning technique to predict outbreaks and epidemics. *International Journal of Advanced Science and Research*, 3, 19–24.
- Soliman, H., Tabak, F., & Sciences, P. (2020). Deep learning framework for RDF and knowledge graphs using fuzzy maps to support medical decision. *Journal of International Research in Medical and Pharmaceutical Sciences*, 14, 92–97.
- Statista. (2019). top-us-social-networking-apps-ranked-by-sessionlength In (Vol. 2020). online.
- Tober, M. J. M. L. A. (2011). PubMed, ScienceDirect, Scopus or Google Scholar-Which is the best search engine for an effective literature research in laser medicine? *Medical Laser Application*, 26, 139–144.
- Velavan, T. P., & Meyer, C. G. (2020). The COVID-19 epidemic. *Tropical Medicine & International Health*, 25, 278–280.
- WHO. (2020). Coronavirus disease (COVID-19): situation report, 162.
- Xu, Z., Zayed, T., Lin, Y.-H., Wang, S., & Li, H. (2020). Relationship between Social Media and ASCE Code of Ethics: Review and case-based discussion. *Journal of Civil Engineering Education*, 146, 05019002.
- Yusof, A. N. M., Muuti, M. Z., Ariffin, L. A., & Tan, M. K. M. (2020). Sharing Information on COVID-19: The ethical challenges in the Malaysian setting. *Asian Bioethics Review*, 1–13.
- Zarrad, A., Jaloud, A., & Alsmadi, I. (2014). In *The evaluation of the public opinion-a case study: Mers-cov infection virus in ksa* (pp. 664–670). IEEE.