MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI231147

Insights on the Future of Digital Health: An Analysis of Twitter Posts of IMIA Fellows

David T MARC[a,1], Karima LALANI[b], Susan H FENTON[b] and Kerryn BUTLER-HENDERSON[c]
[a] *The College of St. Scholastica, Duluth, Minnesota, USA*
[b] *The University of Texas Health Science Center at Houston, Houston, Texas, USA*
[c] *RMIT University, Melbourne, Victoria, Australia*
ORCiD ID: David T. Marc https://orcid.org/0000-0002-6178-5345, Susan H. Fenton https://orcid.org/0000-0003-2943-311X, Karima Lalani https://orcid.org/0000-0002-5982-160X, Kerryn Butler-Henderson https://orcid.org/0000-0002-6082-2108

Abstract. The COVID-19 pandemic has reshaped technology-enhanced services in health and care organizations globally. As the world pivots towards a post-COVID-19 environment, it is essential to examine emerging trends amongst thought leaders in the health information technology sector. This study queried Twitter feeds of IMIA Fellows from 2013 through 2022, utilizing combinations of sentiment analysis, latent dirichlet allocation, and document analysis methods. The results provided a glimpse of positive sentiment year upon year, with the most negative sentiment prevalent in 2020, due to the onset of the pandemic. The findings from this study can be strategically used to analyze emerging trends in digital health, as well as to shape health IT thought leadership in the post-pandemic landscape.

Keywords: COVID-19, sentiment analysis, digital health, latent dirichlet allocation

1. Introduction

The rapid adoption of technology due to the need to digitize services during the coronavirus (COVID-19) pandemic has seen a shift in digital strategies in health and care organizations and in government policy [1-3]. This transformation in digital health has reduced inequities in access to care, whilst increasing the digital divide between the haves and the have-nots [4]. And whilst it could not be predicted just how much the pandemic was going to transform the delivery, support, and management of digital health care, many lessons can be learned to forecast future directions.

At a time when health and care organizations shift into the new normal and re-examine their futures and the strategies required to achieve future goals, it is opportune to examine the emerging paradigms in digital health. The growth in the fifth industrial revolution will have significant impacts on health and care as it is centered on the human-technology interaction or synergistic collaboration between the cyber-physical-biological [5]. Therefore, the healthcare sector needs to be alert and agile in this rapidly shifting ecosystem, and be ready to evaluate and adopt meaningful and safe technologies that support citizen-centric care and promote wellness. This paper reports findings from an

[1] Corresponding Author: David T Marc, email: dmarc@css.edu.

analysis of Twitter posts of digital health thought leaders and influencers to determine the emerging trends in digital health.

2. Methods

Using a document analysis approach, authors KBH and KL downloaded the list of IMIA Fellows (as of 19 October 2022) and searched Twitter, the IMIA profile, and a web profile, where available, for each Fellow. Where a Twitter handle was identified, the number of followers, activity status for the past twelve months, country (where available), public profile status, and the Twitter URL were captured in a spreadsheet. Those with an active, public Twitter profile, and with more than 100 followers were included in the data extraction stage. The final extraction included 48 Twitter profiles. The rationale for selecting Twitter for tracking digital health trends rather than another social media platform was largely based on the accessibility of longitudinal data using the Twitter API.

Using the *rtweet* package in the R statistical programming language [7], the tweets from the Twitter followers were queried. The query included the user's handle, text for the tweet, date, language, retweet count, favorite count, and hashtags. The data was filtered to only include tweets from 1/1/2013 through 11/22/2022, which approximated ten years of data. Retweets were retained in the data for analysis. All tweets were translated into English using Google Translate.

The data was prepared by being tokenized into individual terms, converted to lowercase, and punctuation, numbers, and white space were removed. The frequency of terms was determined across all of the data and the top 20 terms were summarized.

Sentiment analysis was performed using the Bing sentiment lexicon [8] to determine if a word is categorized as positive or negative. The sentiment analysis was performed for the entire data corpus as well as for the data corpus broken down by year.

To perform a topical analysis, a simple text corpus was created and prepared for analysis using the *tm* package [9]. The tweets were prepared by transforming the data to lowercase text, removing English stopwords, removing punctuation, removing numbers, and removing white space. A document term matrix was computed from the processed text corpus to obtain the word frequency of each tweet. Words were only included when they occurred at least 5 times and had a minimum number of 3 characters. Empty rows were also removed from the document term matrix.

Latent Dirichlet Allocation (LDA) was performed on the document term matrix to determine the topics found in the tweets. LDA is a probabilistic model that can be applied to a corpus of text from documents with a random mixture of latent topics, where each topic is characterized by a distribution of words [10]. LDA can identify topics based on word frequencies that co-occur in documents. An LDA analysis was performed separately for each year of Twitter data to obtain the change in topics over time. The optimal number of topics was determined using the methods described by Cao Juan [11].

3. Results

There were a total of 82,001 unique tweets from 48 Twitter accounts. 36,313 of the unique tweets were retweets. The remaining 46,688 original tweets were shared 131,324 times and favored 484,368 times. There were a total of 39 languages detected in the tweets, the most common being English (n= 70,813), Arabic (n=3,607), Spanish (n=3,444), and

Finnish (n=809). The most common hashtags included COVID19 (n=5,646), hcsm (n=4,093), hcsmanz (n=3,073), AI (n=2,827), and digitalhealth (n=2,656). The most common words included health (n=9,760), data (n=4,738), informatics (n=3,594), covid (n=3,452), and care (n=3,157).

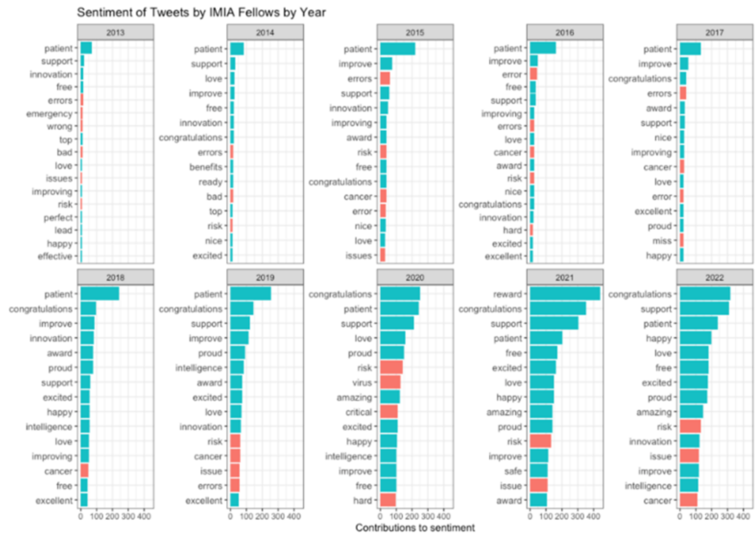

Figure 1. The top 20 negative and positive terms from 2013 through 2022 tweeted by IMIA Fellows.

Figure 1 compares the sentiment of the top 20 terms from 2013 through 2022. There are some notable observations in the data. The term patient has a positive sentiment and was the top term from 2013 through 2019. While the top term in 2020 and 2022 was congratulations with a positive sentiment. In 2021, the top term was reward with a positive sentiment. In 2020, there was the greatest use of terms with a negative sentiment with the most common negative terms including risk, virus, and hard. Prior to 2021, cancer and error were the most common negative terms used in tweets. In 2021 and 2022, risk and issue were the most common negative terms.

As shown in Table 1, when examining the topics emerging from each of the years of Twitter data, the number of topics ranged from 3 to 8. In 2013 and 2014, Health and HealthIT emerged as common topics of tweets. In 2015 and 2016, topics related to big data, AMIA, safety, and EHR were founds. In 2017, digital health and MedInfo emerged. In 2018, a topic related to thanks and congratulation emerged, and included a topic on patient safety and the EHR. In 2019, topics were related to patient safety, digital health, MedInfo, and AMIA. In 2020, COVID emerged as a new topic and was covered in context to people, information, and research. In 2021, COVID was now related to vaccines, and topics emerged related to God, research, people, clinical care, and time. In 2022, the most common topic was congratulatory related to AMIA, as well as COVID related to time, research, God, and openEHR.

Table 1. Results of the LDA Topical Analysis from 2013-2022 from IMIA fellow tweets.

Year	Frequency	Proportion of Tweets For Each Topics	
2022	Docs: 18,553 Terms: 6,291	0.15: informatics great health 0.15: covid people just now time 0.14: health new acijournal care data 0.14: health research will university	0.14: god israeli palestine people 0.14: data can need openehr will 0.13: atari one team first day
2021	Docs: 18,600 Terms: 6,320	0.17: just like one year last 0.16: data health new care clinical 0.15: people can will many world 0.14: amia congratulations informatics	0.14: health research science digital 0.14: covid vaccine people pandemic 0.09: may god reward kuwait allah
2020	Docs: 13,992 Terms: 5,427	0.23: just one good like great 0.21: informatics health amia great 0.20: covid people coronavirus care	0.18: health data covid information 0.17: covid new care health paper
2019	Docs: 7,727 Terms: 3,268	0.28: can just one will get 0.26: care new health patient safety	0.24: informatics great medinfo 0.22: health data information digital
2018	Docs: 6,137 Terms: 2,653	0.18: will can just like make 0.17: paper new safety patient ehr 0.17: health care data patient hcsm	0.17: great hic thanks digitalhealth 0.16: informatics health amia data 0.15: health efmi mie informatics
2017	Docs: 3,413 Terms: 1,404	0.24: safety patient care hcsm hitsm 0.20: hercfarr data chcnorth health 0.19: infomedhiba medinfo conarpe	0.19: amia informatics great 0.18: health informatics digital data
2016	Docs: 3,511 Terms: 1,422	0.15: just will thanks time acr 0.15: informatics amia amiainformatics 0.14: health hcsm care hitsm healthit 0.14: patient safety ehr new	0.14: via can data privacy one 0.14: hercfarr data health https 0.13: data health healthcare efmi
2015	Docs: 5,404 Terms: 2,021	0.14: ehr safety patient health 0.13: can people will need via 0.13: informatics amia health 0.13: great openmrs thanks day see	0.12: hcsm hitsm health care hcldr 0.12: data via big privacy medical 0.12: health care healthit healthcare 0.12: mahealth endocrinewitch health
2014	Docs: 2,748 Terms: 1,043	0.36: healthxph acr good endocrinewitch 0.32: informatics amia data	0.32: health hcsmanz care hitsm
2013	Docs: 1,675 Terms: 603	0.37: endocrinewitch openmrs will one 0.33: via zite health healthit care	0.30: hcsmanz health care informatics

COVID-19 was a major disruptor to healthcare, which was found in the Twitter feeds particularly in 2020. That year, there was the greatest negative sentiment compared to the other years. This sentiment aligned with the topics that emerged including the coverage of COVID-19 related to people, information, and research. In 2021, the sentiment was more positive with the terms reward, congratulations, support, and patient commonly occurring while, the negative terms were related to risk and issues. Finally, in 2022, the most common topic was a congratulatory topic related to AMIA. In 2021 and 2022 COVID-19 was also a common topic related to time, research, God, and openEHR.

The results of this study established that longitudinal Tweets from IMIA Fellows offer insight into the major topics and events drawing attention in health IT-related professions, and in healthcare more generally. Other studies found Twitter to be a useful tool for discovering healthcare trends, particularly related to COVID-19 [12-14]. There have been several articles that investigate the use of Twitter to disseminate information related to COVID-19 and recommend that the information be trusted cautiously [15]. This same article suggested that the information on Twitter can be used to inspire audiences to act or investigate a topic on their own by validating what they are reading with trusted sources.

5. Conclusions

Tweets of IMIA fellows were queried from 2013 to 2022. The topics covered during this period focused on patient safety, errors, digital health, data, professional gatherings (i.e., MedInfo), and COVID-19. The sentiment in tweets was rather positive, with the most negative sentiment being present in 2020 related to virus and risks. Given the spread of COVID-19 in 2020, the emergence of these terms is not surprising.

COVID-19 was a significant event that influenced conversations happening on Twitter including the posts by influential IMIA Fellows. Future work could use Twitter data to understand critical topics that health IT professionals are concerned with globally.

References

[1] Pérez Sust P, Solans O, Fajardo JC, Medina Peralta M, Rodenas P, Gabaldà J, Garcia Eroles L, Comella A, Velasco Muñoz C, Sallent Ribes J, Roma Monfa R, Piera-Jimenez J. Turning the Crisis Into an Opportunity: Digital Health Strategies Deployed During the COVID-19 Outbreak. JMIR Public Health Surveill. 2020 May;6(2):e19106, doi: 10.2196/19106.

[2] Kadakia K, Patel B, Shah A. Advancing digital health: FDA innovation during COVID-19. NPJ Digit Med. 2020 Dec;3(1):161, doi: 10.1038/s41746-020-00371-7.

[3] Peek N, Sujan M, Scott P. Digital health and care in pandemic times: impact of COVID-19. BMJ Health Care Inform. 2020 Jun;27(1):e100166, doi: 10.1136/bmjhci-2020-100166.

[4] Crawford A, Serhal E. Digital Health Equity and COVID-19: The Innovation Curve Cannot Reinforce the Social Gradient of Health. J Med Internet Res. 2020 Jun;22(6):e19361, doi: 10.2196/19361.

[5] Noble SM, Mende M, Grewal D, Parasuraman A. The Fifth Industrial Revolution: How harmonious human–machine collaboration is triggering a retail and service [r] evolution. J Retail. 2022 Jun;98(2):199-208, doi: 10.1016/j.jretai.2022.04.003.

[6] Kearney MW. rtweet: Collecting and analyzing Twitter data. J Open Source Softw. 2019 Oct;4(42):1829, doi: 10.21105/joss.01829.

[7] R Core Team. R: A language and environment for statistical computing. Stat Comput Vienna Austria, Vienna, Austria. 2021.

[8] Hu M, Liu B. Mining and summarizing customer reviews. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining 2004 Aug 22 (pp. 168-177), doi: 10.1145/1014052.1014073.

[9] Feinerer I, Hornik K. tm: Text Mining Package. R package version 0.7-9. 2022.

[10] Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. J Mach Learn Res. 2003;3(Jan):993-1022.

[11] Cao J, Xia T, Li J, Zhang Y, Tang S. A density-based method for adaptive LDA model selection. Neurocomputing. 2009 Mar;72(7-9):1775-81, doi: 10.1016/j.neucom.2008.06.011.

[12] Slobodin O, Plochotnikov I, Cohen IC, Elyashar A, Cohen O, Puzis R. Global and Local Trends Affecting the Experience of US and UK Healthcare Professionals during COVID-19: Twitter Text Analysis. Int. J Environ Res Public Health. 2022 Jun;19(11):6895, doi: 10.3390/ijerph19116895.

[13] Nakagawa K, Yang NT, Wilson M, Yellowlees P. Twitter Usage Among Physicians From 2016 to 2020: Algorithm Development and Longitudinal Analysis Study. J Med Internet Res. 2022 Sep;24(9):e37752, doi:10.2196/37752.

[14] Saura JR, Ribeiro-Soriano D, Saldaña PZ. Exploring the challenges of remote work on Twitter users' sentiments: From digital technology development to a post-pandemic era. J Bus Res. 2022 Mar;142:242-54, doi: 10.1016/j.jbusres.2021.12.052.

[15] Rosenberg H, Syed S, Rezaie S. The Twitter pandemic: The critical role of Twitter in the dissemination of medical information and misinformation during the COVID-19 pandemic. Can J Emerg Med. 2020 Jul;22(4):418-21, doi: 10.1017/cem.2020.361.