JMIR Publications
The Leading eHealth Publisher

Perception of Online Communities towards the Use of Cancer Immunotherapy: A Data Mining Study of 3.6 Million Web-based Posts from Social Media Platforms Using BERTopic

Xingyue Wu, Chun Sing Lam, Ka Ho Hui, Herbert Ho-fung Loong, Rui Zhou, Chun-kit Ngan, Yin Ting Cheung

Submitted to: Journal of Medical Internet Research
on: May 27, 2024

Preprint
JMIR Publications

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on it's website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressively prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript **4**
Supplementary Files **19**
Figures **20**
Figure 1 ..... **21**
Figure 2 ..... **22**
Multimedia Appendixes **23**
Multimedia Appendix 1 ..... **24**
Multimedia Appendix 2 ..... **24**
Multimedia Appendix 3 ..... **24**
Multimedia Appendix 4 ..... **24**

Preprint JMIR Publications

Perception of Online Communities towards the Use of Cancer Immunotherapy: A Data Mining Study of 3.6 Million Web-based Posts from Social Media Platforms Using BERTopic

Xingyue Wu[1*]; Chun Sing Lam[2*]; Ka Ho Hui[2]; Herbert Ho-fung Loong[3]; Rui Zhou[2]; Chun-kit Ngan[4]; Yin Ting Cheung[5]

[1]School of Pharmacy, Faculty of Medicine The Chinese University of Hong Kong The Chinese University of Hong Kong Hong Kong HK
[2]School of Pharmacy, Faculty of Medicine The Chinese University of Hong Kong Shatin, Hong Kong SAR, China Hong Kong HK
[3]Department of Clinical Oncology, Faculty of Medicine The Chinese University of Hong Kong Shatin, Hong Kong SAR, China Hong Kong HK
[4]Data Science Program Worcester Polytechnic Institute Worcester, MA, United States. Worcester US
[*]these authors contributed equally

Corresponding Author:
Yin Ting Cheung

Abstract

Immunotherapy has become a game changer in cancer treatment. Few studies have investigated the perceptions about its use by analyzing social media data. This study aimed to use a topic modeling technique, BERTopic, to explore the perceptions of the online cancer communities regarding immunotherapy. A total of 4.9 million posts were extracted and preprocessed. BERTopic modeling was performed to identify topics from the posts, which were then broadly grouped into distinct themes. 3.6 million posts were remained for modeling after data cleaning. The highest overall topic quality achieved by BERTopic was 70.47% (topic diversity: 87.76%; topic coherence: 80.21%). BERTopic generated 14 topics related to the perceptions of immunotherapy, which were categorized into six themes. The themes primarily covered (1) hopeful prospects offered by immunotherapy, (2) perceived effectiveness of immunotherapy, (3) complementary therapies or self-treatments, (4) financial and mental impact of undergoing immunotherapy, (5) impact on lifestyle and time schedules, and (6) side effects due to treatment. This study provided an overview of the multifaceted considerations essential for the application of immunotherapy as a therapeutic intervention. Furthermore, it demonstrated the effectiveness of BERTopic in analyzing large amounts of data to identify perceptions underlying social media and online communities.

(JMIR Preprints 27/05/2024:60948)
DOI: https://doi.org/10.2196/preprints.60948

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Preprint
JMIR Publications

Original Paper

Xingyue Wu[a#], Chun Sing Lam[a#], Matthew Ka Ho Hui[a], Herbert Ho-fung Loong[b], Keary Rui Zhou[a], Chun-kit Ngan[c], Yin Ting Cheung[a*]
[a] School of Pharmacy, Faculty of Medicine, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China.
[b] Department of Clinical Oncology, Faculty of Medicine, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China.
[c] Data Science Program, Worcester Polytechnic Institute, Worcester, MA, United States.
[#] These authors contributed equally to this work.

*Corresponding Author:
Yin Ting Cheung, PhD
Associate Professor
School of Pharmacy, Faculty of Medicine
The Chinese University of Hong Kong
8th Floor, Lo Kwee-Seong Integrated Biomedical Sciences Building Area 39
The Chinese University of Hong Kong, Shatin, N.T.
Hong Kong SAR, China
Phone: 852 3943 6833
Fax: 852 2603 5295
Email: yinting.cheung@cuhk.edu.hk

Preprint
JMIR Publications

Perception of Online Communities towards the Use of Cancer Immunotherapy: A Data Mining Study of 3.6 Million Web-based Posts from Social Media Platforms Using BERTopic

Abstract

Immunotherapy has become a game changer in cancer treatment. Few studies have investigated the perceptions about its use by analyzing social media data. This study aimed to use a topic modeling technique, BERTopic, to explore the perceptions of the online cancer communities regarding immunotherapy. A total of 4.9 million posts were extracted and preprocessed. BERTopic modeling was performed to identify topics from the posts, which were then broadly grouped into distinct themes. 3.6 million posts were remained for modeling after data cleaning. The highest overall topic quality achieved by BERTopic was 70.47% (topic diversity: 87.76%; topic coherence: 80.21%). BERTopic generated 14 topics related to the perceptions of immunotherapy, which were categorized into six themes. The themes primarily covered (1) hopeful prospects offered by immunotherapy, (2) perceived effectiveness of immunotherapy, (3) complementary therapies or self-treatments, (4) financial and mental impact of undergoing immunotherapy, (5) impact on lifestyle and time schedules, and (6) side effects due to treatment. This study provided an overview of the multifaceted considerations essential for the application of immunotherapy as a therapeutic intervention. Furthermore, it demonstrated the effectiveness of BERTopic in analyzing large amounts of data to identify perceptions underlying social media and online communities.

Keywords: social media; cancer; immunotherapy; perceptions; data mining

Preprint JMIR Publications

1. Introduction

In recent years, immunotherapy has emerged as one of the most promising therapeutic approaches for treating cancer. It works by activating the innate immune system to identify and attack cancer cells. Immunotherapy encompasses various strategies, including immune checkpoint inhibitors, T-cell transfer therapy, monoclonal antibodies, treatment vaccines, and immune system modulators [1]. These immunotherapeutic strategies have received approval for the treatment of several cancer types, such as lung cancer, prostate cancer, chronic lymphocytic leukemia, and non-Hodgkin's lymphoma [2]. Immunotherapy offers substantial benefits in terms of precision, specificity, and long-term survival improvements, representing a significant breakthrough in cancer treatment [3].

Immunotherapy has made remarkable progress and demonstrated clinical value. However, there are drawbacks that may hinder its clinical use and acceptance by patients with cancer. One notable limitation is the considerable variability in individuals' responses to immunotherapy. While the treatment may be effective for some patients with a specific type of cancer, it may not be as effective for others with the same cancer type [4]. Besides, immunotherapy-related adverse events (irAEs) have been observed with the increasing frequency and duration of immunotherapy usage. Patients' decisions to undergo immunotherapy are influenced by a range of factors, including their perceptions of its efficacy, side effects, procedural aspects, costs, their levels of knowledge about the treatment, and the comprehensiveness of advice provided by healthcare providers [5; 6]. These findings highlight the importance of understanding patients' perspectives regarding immunotherapy to manage the uncertainties faced by patients and make informed decisions whether to proceed with immunotherapy.

The Internet has become an indispensable source of health information for patients [7]. Many patients scour the Internet for medical guidance, share their personal experiences, and interact on social media platforms [8]. The Internet, therefore, provides a valuable platform for capturing diverse perspectives. Several studies have explored these perspectives by collecting data from social media platforms and forums [9-13]. For example, through analyzing posts from social media, a study revealed that pain and fatigue were the most commonly discussed symptoms among patients with non-small-cell lung cancer regarding the use of immunotherapy [10]. Another study found that 55% of patients' posts and 37% of caregivers' posts expressed positive perceptions of immunotherapy for treating advanced bladder cancer [11]. Notably, these previous studies usually extracted posts from a single site and analyzed only a small portion of the available posts. Moreover, these studies primarily focused on patients with specific cancer types, and thus the findings may not be generalizable to patients with other cancer types for which immunotherapy has been approved [13].

One challenge in analyzing data from the Internet and social media posts is dealing with vast amounts of unstructured text, which necessitates extensive preprocessing measures prior to analysis. Different machine-learning techniques have been explored to increase the efficiency of online text analysis. BERTopic, a topic modeling technique that is an extension of the machine learning model––Bidirectional Encoder Representation from Transformers (BERT)––has been used to identify the underlying public perceptions within social media posts. It has demonstrated the capability to unveil latent patterns and extract topics from large datasets [14-16]. Compared to other models, BERT models provide a stronger understanding of the contextual meaning of each word in document representations, which is attributable to their bidirectional training of transformers [17]. This enables the generation of more accurate topics than those generated using traditional statistical models. Moreover, the BERT model is continuously trained and updated by researchers, ensuring ongoing improvements [18].

However, no study has employed this technique to explore the viewpoints of the online cancer community on immunotherapy.

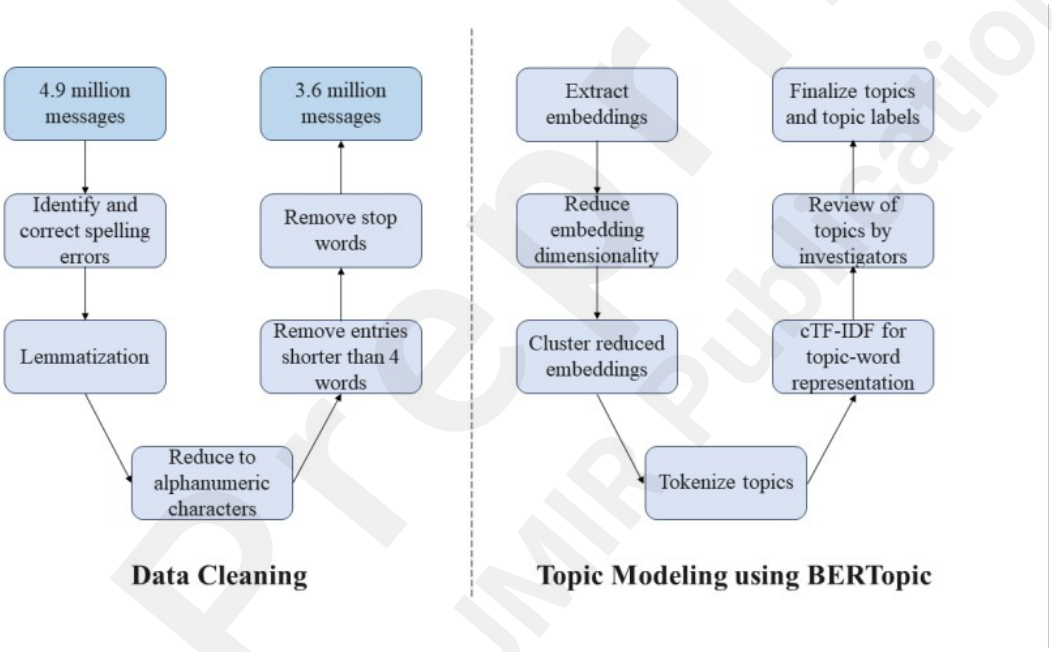
With the objective of understanding the perspectives of online cancer communities regarding immunotherapy, this study used BERTopic to analyze a large number of posts extracted from multiple social media and online forums. The findings of this study may offer valuable insights for clinicians, patients, and researchers to enhance decision-making processes when considering immunotherapy as a treatment option.

2. Methods

2.1 Overview

This was a retrospective study that involved analyzing texts collected from social media platforms and online forums. The study proceeded through four primary methodological phases: (1) collecting the textual data, (2) cleaning and processing the extracted texts, (3) performing and optimizing topic modeling, and (4) conducting thematic analysis. The study process workflow is illustrated in Figure 1. Ethical approval was exempted, as the study exclusively utilized anonymized data from publicly available sources.

Figure 1. Flowchart depicting the data mining process.



2.2 Data sources and collection

To gather data for this study, textual data were collected from online forums and social media platforms. To ensure consistency and minimize translation concerns, only English-language texts were included. Posts were extracted from three prominent social media platforms: Facebook, Twitter, and Reddit. These platforms were chosen for their popularity and the substantial textual content they contain [19]. Sixteen different online cancer-related forums were also selected for analysis. The posts from these platforms dated before November 15, 2022 were included, with varying start dates (Supplementary Table 1). The selection criteria for these cancer-related online forums have been detailed in a previous study [20]. Search keywords related to immunotherapy were identified from multiple authoritative websites, including the American Cancer Society, Cancer Research UK, and the National Cancer Institute (Supplementary Table 2). Site-provided search engines were utilized to identify related posts.

2.3 Data pre-processing

Spark NLP, a natural language processing library for Python built on top of Apache Spark, was

utilized to preprocess the textual data [21]. The texts extracted from the forums and social media sites were subjected to a thorough cleaning process to enhance their suitability for further analysis. The cleaning process involved multiple steps, including checking and correcting spelling errors, lemmatization, removal of stop words, and elimination of short entries (Figure 1). Each text message was first spellchecked and the vocabulary was reduced to its base root form using Spark NLP. After that, each word vocabulary was reduced to alphanumeric characters that contained only letters from “a to z”, “A to Z” and “0 to 9”. To generate a precise and semantically meaningful text corpus, any message composed of fewer than four words, as well as commonly occurring words (e.g., the, a, and, and in) that carried little or no meaning in the message, were removed. Finally, the cleaned text messages were processed using the medical language models from Healthcare Spark NLP, a pretrained pipeline to recognize medical terminologies within the messages [22].

2.4 Topic modeling

The next step involved the extraction of topics from the data. Topic modeling, an unsupervised machine-learning approach, was employed to unveil hidden semantic structures and extract distinct topics from within the extracted textual data [23]. In this study, we used BERTopic, a topic modeling library that harnesses the vectors generated by BERT models. Using BERTopic, all texts were passed through the BERT model to create embeddings that converted text messages into numerical representations. The dimensions of the embedding vectors were reduced due to the “Curse of Dimensionality” [24]. The bisecting K-means algorithm was used to group the reduced-dimension embedding vectors into clusters until optimal silhouette scores were achieved [25]. Various numbers of topics, ranging from 5 to 25, were tested to determine the optimal metrics. Topic diversity (indicating the model’s ability to capture differences between generated topics), coherence (the frequency of descriptive words of the topic within each cluster), and quality (topic diversity multiplied by topic coherence) were used as metrics to evaluate the effectiveness of isolating topics from the texts [26; 27].

Following the identification of clusters with the highest topic quality, the 20 most frequent words from each topic were extracted. BERTopic modeling and topic generation were conducted using the Worcester Polytechnic Institute’s Turing cluster, which has high computing power for large volumes of data processing and is located at Worcester Polytechnic Institute.

2.5 Sentiment and Thematic analysis

Sentiment analysis was performed using two popular libraries, VADER (Valence Aware Dictionary and sEntiment Reasoner) and Textblob [28; 29]. Both libraries were used to enhance the accuracy of the results and minimize potential biases that may arise from relying on a single library. The sentiment scores obtained from each library were then averaged to derive a final sentiment score, which ranged from -1 (indicating a very negative sentiment) to 1 (indicating a very positive sentiment). Based on the topics generated through BERTopic modeling and the assigned labels, thematic analyses were conducted to identify themes associated with immunotherapy. Each theme could correspond to one or more topics, as topics with similar labels were grouped together. In the initial round of thematic analysis, four domain experts (including oncologists and oncology pharmacists) were provided with the keywords related to the topics, as well as representative posts for each topic. Any themes that achieved consensus among all domain experts were selected, while themes without consensus and any discrepancies were documented and presented for discussions among the domain experts in the next round. Only the themes that were agreed by 75% of the domain experts in either Round 1 or Round 2 were summarized.

3. Results

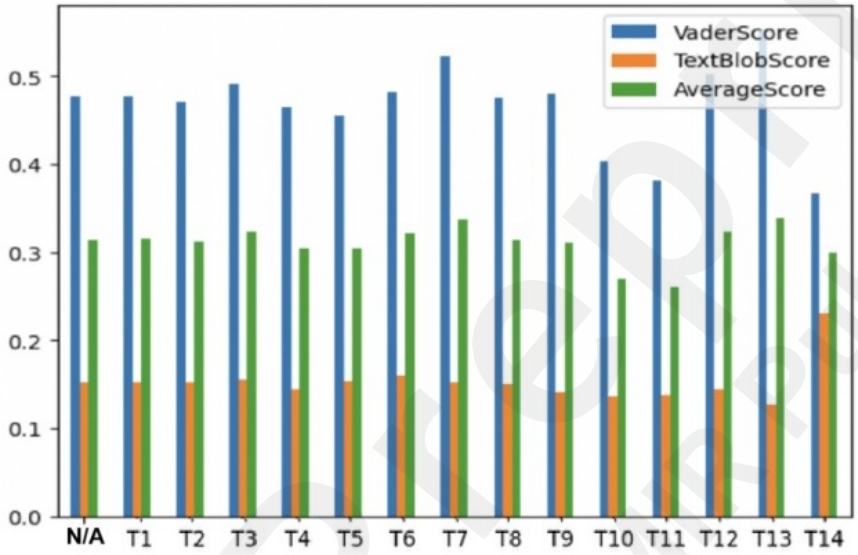
3.1 Overview

A total of 4.9 million posts were gathered from online forums and social media platforms. After data cleaning, 3.6 million posts remained for further analysis (Figure 1).

3.2 Number and qualities of topics

Supplementary Table 3 presents BERTopic's performance across different numbers of topics to determine the optimal value. Fifteen topics achieved the highest overall topic quality, scoring 70.47% (with the highest topic diversity at 87.76% and a high topic coherence of 80.21%). We also used these metrics to assess the quality of individual topics generated by the model, yielding topic quality scores ranged from 78% to 88% (Supplementary Figure 1). One topic (Topic 15) was excluded due to the mismatch between its extracted words and the topics related to immunotherapy. Based on the sentiment analysis results, the average sentiment score of around 0.3 across the 14 topics preliminarily suggests a generally positive sentiments towards immunotherapy within the online community (Figure 2). A more comprehensive interpretation of the specific topics is presented in the subsequent sections.

Figure 2. Sentiment analysis results.



3.3 Themes and topics

A total of six themes have been identified (Table 1). Theme 1 centered on the feelings of hope and positivity among patients treated with immunotherapy. This theme highlighted the belief that immunotherapy offers hope for patients with cancer to lead normal lives and provides a potential cure for cancer with minimal side effects. Theme 2 described immunotherapy as a more effective treatment option after other treatments have failed. Patients may have tried other cancer treatments, such as chemotherapy and radiotherapy, but only found successful results with immunotherapy. Besides, patients believe combining immunotherapy with other conventional therapies, particularly chemotherapy, can lead to higher efficacy in treating cancer than using a single treatment approach. Theme 3 suggested that patients receiving immunotherapy also explore diets or complementary medicines to manage their symptoms or side effects from treatment or the cancer itself. For example, patients may employ mind–body practices, such as massage and meditation, or use herbal medicines to alleviate certain physical and psychological side effects caused by immunotherapy. Probiotics were frequently mentioned by users as a potential way to boost the effectiveness of immunotherapy.

In Theme 4, the potential ineffectiveness, worse outcomes, and high cost of treatment

were frequently mentioned among patients, which acted as hindrances to the use of immunotherapy. Theme 5 suggested that immunotherapy can disrupt patients' normal daily lives, particularly for those enrolled in immunotherapy trials. Additionally, the treatment may negatively impact their activities of daily living, such as taste changes, dietary restrictions, and time consumption. Theme 6 indicated that immunotherapy may cause physical and psychological symptoms and side effects, like rash, inflammation, and nephrotoxicity. These adverse effects may sometimes pose greater challenges compared to those arising from other treatments or the cancer itself. Some patients acknowledged that these symptoms may be unavoidable alongside the curative effects of immunotherapy.

Table 1. Themes and selected posts under each theme.

Topics and related keywords	Example of quotes
Theme 1: Hopeful prospects offered by immunotherapy	
Topic 1: good, time, hope, love, great, feel, work, lot, treatment, start	• *"...it does help to not feel so alone with this as he was diagnosed this time last year and since then it has been a roller coaster my husband is on immunotherapy already and it stopped the growth quickly i hope it works as well for your husband."* • *"I only can tell you he is doing very well and thru prayer and this new treatment i feel we all have a new found hope"*
Topic 3: fight, hope, positive, news, treatment, glad, beat, good, feel, life	• *"He was able to have immunotherapy for his stage 3b non small cell lung cancer via a clinical trial and has also had positive results. his trial has finished now but like yourself, would like to continue with this treatment..."* • *" I have heard very positive things about immunotherapy. in my mind, it seems logical to use your own immune system to deal with cancer cells - sounds less damaging than chemo or radiotherapy"*
Topic 5: hope, day, love, morning, huge, satisfy, glad, feel, good, love	• *"We can only but hope the researchers and doctors one day are able to cure all cancers without side effects. I think the answer may well be in developing Immunotherapy tailored to the individual."* • *"The day will come, hopefully soon, when immunotherapy will be available to all. That is the only way that our unique cancer cells will be eradicated. It can't come soon enough."*
Topic 6. saint, faith, family, love, god, strength, miracle, wonderful, hope, glad	• *"The drug has been a miracle for her and she has responded really well to the drug essentially keeping the cancer at bay."* • *"there absolutely are some immunotherapies that are working. i know someone who was stage 4, with very advanced disease from what i understand, that is now ned due to opdivo... it won't help everyone. however, for those it does help, it is a miracle."*
Theme 2: Perceived effectiveness of immunotherapy	
Topic 4. gallbladder, liver, bile, symptom, surgery, scan, remove, problem, ultrasound, test	• *"i was given the chance to try nivolumab an immunotherapy drug as one last hail mary pass. unbelievably i have had a tremendous response, with ky abdominal lymph nodes, adrenal glands lungs and liver now clear."*

	• *"he has had 3 rounds of this immunotherapy drug and is due ct scan tmrw to see how it's working. he is convinced the treatment is not working as he has started having swallowing difficulty again which is only symptom he initially had."*
Topic 13. risk, study, benefit, favorable, evidence, effect, patient, data, high, treat	• *"I had 6 weeks of daily radiotherapy and 6 weekly courses of Chemotherapy. This proved completely ineffective. I am now on Immunotherapy coming up to the 8th round. This does appear to have been successful in that the latest CT scan compared to one I had in June shows the cancer has shrunk..."* • *"Congratulations on having a favorable response against the melanoma. Fifteen years ago, metastatic melanoma usually meant your family would be making a large payment to a funeral home in the future. immunotherapy has given us our lives back"*
Topic 8. antioxidant, radiation, chemo, protect, treatment, dose, effect, radical, kill, interfere	• *"Better results with less side effects... combining chemo with immunotherapy drugs gives the potential for long term remission and even a cure for advanced cancers."* • *"... chemotherapy and immunotherapy is often very effective in treating liver tumours and many patients livers become operable."*
Theme 3: Complementary therapies or self-treatments	
Topic 2. acupuncture, pain, feel, needle, massage, neuropathy, flash, hot, session, side	• *"...when i was on ac chemo my immune system was bad so onc said no acupuncture since that might increase my risk of getting sick. now i'm on immunotherapy and she is encouraging me to get massage and acupuncture."* • *"I've had two sessions of acupuncture to try to help with multiple side effects of immunotherapy (abdominal pain, constipation, nausea, reflux, severe neuropathy, muscle pain and bilateral bell's palsy)"*
Topic 10. keto, diet, food, organic, probiotic, supplement, eat, weight, sugar, body	• *"Immunotherapy in turn is very dependent on good gut bacteria and strongly stimulated by probiotics that have Bifidobacterium..., but for anyone taking immunotherapy its probably a good idea to add probiotics as well."* • *"Have you tried anything else, e.g., diet, nutrition, supplements etc? the oncologist won't advise on any of it and we don't have 'integrated treatment' in nz at all so i have been looking into it myself and researching it as much as possible, keen to do anything we can to give best outcome!"*
Theme 4: Financial and mental impact of immunotherapy	
Topic 7. diagnosis, feel, anxiety, traumatic, stress, cost, experience, time, depression, therapist	• *"It is difficult to sum it up but I do not feel that optimistic about it (immunotherapy). Everything seems to be about expensive drugs with nasty side effects."* • *"But unfortunately I'm living in the poor country, the government and insurance didn't pay for my treatment. And immunotherapy costs a lot that I can't afford it."* • *"dealing with progression is profoundly stressful, and after first or second-line... immunotherapy isn't seemingly moving*

	at all for er-positive cancers..." • *"She is due to start immunotherapy pembrolizumab on april 13th. the long wait for treatment has been horrendous as each day she seems to be getting more depressed and getting weaker."*
Theme 5: Impact on lifestyle and time schedules	
Topic 9. simulation, schedule, treatment, appointment, ill, week, start, time, plan, session	• *"Apparently I am eligible to be included in a trial of immunotherapy which would occur after my chemoradiation therapy and before surgery. It would involve an eight week program and doesn't sound at all nice, especially after the previous treatments..."* • *"Ideally I'd like to have taken a break and/or reduce dosages rather than quit altogether especially since we spent 5 weeks to decide on a new treatment plan rather than have one at hand... But since this was a trial I think they don't have too much flexibility."*
Theme 6: Side effects due to treatment	
Topic 11. ill, swell, weak, surgery, compression, feel, wear, inflame, tight, back	• *"Following the four rounds of nivolumab and ipilimuab i had six weeks off. the side effects were quite minor and wore off pretty quickly. (skin rash, muscle aches, headaches, fatigue and general cold/flu symptoms).."* • *"I was a bit surprised at how the side effects returned so quickly... I couldn't breath so presented to the hospital emergency department. ct scans and a chest x-ray showed swelling but little else... the issue was diagnosed as pneumonitis (inflammation in the lungs). ..."* • *"The current treatment of Keytruda, Carboplatin and Gemzar is keeping my fast growing tumors stable. But the immunotherapy has resulted in inflammation of lungs so I may have to stop it for now... I don't know how long it will take to resolve this lung issue... Sometimes the treatments can result in worst side effects than the disease itself."*
Topic 12. immune, damage, skin, fever, burn, kill, system, therapy, tissue, body	• *"Apparently I have developed an acute kidney injury. Apparently it is a drug induced autoimmune thing....so I guess I was responding to the immunotherapy! ... I think they caught it fairly early and am hopeful that things can be restarted while closely monitored."* • *"long story short i have been diagnosed with type2 diabetes believed to be at least partially due to immunotherapy (auto immune reaction)...i am having a very hard time with this..."* • *"I've heard rashes are actually good with some immunotherapy meds, it shows that it's working....except not necessarily for keytruda which I'm on. I've had several other side effects, but not a rash, and I have had a good response."*
Topic 14. pain, symptom, hope, painful, colitis, arthritis, response, issue, experience, costochondritis	• *"sorry too, about your back pain. have you seen it mentioned as a side effect of the immunotherapy at all? hoping you get some relief or at least some acknowledgement soon."* • *"for now kadcyla is working for me, but i do have a rash on the back of my neck and serious neuropathy. i've found a*

	cream that helps with it, along with pain meds i take at bedtime."

4. Discussion

We analyzed a large dataset comprising 3.6 million posts from three social media platforms and 16 online cancer forums, providing a comprehensive overview of perceptions regarding immunotherapy within English-speaking online cancer communities. A cutting-edge natural language processing technique, BERTopic, was used to generate interpretable topics, ensuring robust data analysis. The identification of different themes underscores the diverse attitudes toward immunotherapy across various aspects. These themes may also offer valuable insights into the facilitators and barriers to the adoption of immunotherapy. In addition, the results showed that using machine-learning techniques to analyze patient-generated online textual data can enhance our understanding of perspectives on newer cancer treatments, and such information may facilitate communication between patients and clinicians when discussing immunotherapy as a therapeutic option.

The positive factors that encourage the utilization of immunotherapy included its comparatively mild side effect profile and the potential to instill hope in patients by offering a viable treatment option when other methods have failed. Our findings also indicated that online cancer communities seem to be well-informed, with some patients well-versed in the benefits of immunotherapy. Laypeople now access medical information on the Internet independently and acquire firsthand information about advances in cancer treatment. Access to online health information assists patients in making health related decisions, leading to more professional consultation, improved treatment compliance, and better self-care [30]. However, the Internet should not be viewed as a replacement for professional health information sources [31]. The guidance and advice of healthcare professionals remain indispensable in helping patients make complex medical decisions [32]. Online health information cannot provide detailed diagnoses or personalized treatment plans for patients; moreover, the presence of misinformation on websites might mislead patients about treatment [33]. Taken together, our results highlight the rising health literacy of online cancer communities, especially those involving younger generations. Future efforts should focus on enhancing the accuracy of online health information using reliable rating tools, effective search engine ranking, and progress in crowdsourcing websites.

Although immunotherapy has shown clinical benefits in several trials, there are still potential barriers to its utilization. The high cost associated with immunotherapy emerges as a frequently discussed concern among patients with cancer, emphasizing the need for revising drug valuations and reimbursement models [34]. Many quotes were related to patients' descriptions of their physical symptoms induced by immunotherapy and how it has impeded their activities of daily living and disrupted their normal lives. In addition to physical irAEs, the long-term use of immunotherapy has also been associated with psychiatric side effects such as fatigue, insomnia, anxiety, and depression [35; 36]. Hence, it is important for oncology practitioners to manage patients' expectations and communicate with patients regarding the potential side effects of immunotherapy before initiating treatment. This proactive approach allows patients to better understand the benefits and risks of the treatment, which may prevent them from discontinuing treatment because of known adverse reactions.

Several clinical practice guidelines have been proposed to address the evaluation and management of immunotherapy-induced toxicity [37], and the diagnosis and treatment of toxicity associated with immune checkpoint inhibitor therapy in specific organ systems [38]. The provision of high-quality supportive care from a multidisciplinary team of healthcare professionals can help identify and manage irAEs [39]. In practice, patients may turn to self-

management strategies to cope with these symptoms. We found that some patients mentioned in online posts that they used traditional, complementary, and integrative medicine modalities, such as probiotics, herbal remedies, and meditation, to self-manage the toxicity induced by immunotherapy. Currently, there have been some studies that evaluated how complementary medicines may enhance the efficacy of immunotherapy and alleviate immune toxicity [40; 41]. For example, the treatment of irAEs often necessitates the long-term use of high-dose corticosteroids [42]. Integrating complementary modalities into the management of irAEs may reduce corticosteroid use. Our group has previously reported that patients with cancer frequently seek advice on the use of complementary medicines on social media platforms [20]. At this juncture, clinicians should focus on initiating effective communication regarding the use of complementary medicines and help patients establish realistic expectations while being well-informed about the limited evidence supporting these approaches.

Our study has several limitations. First, the data used in this research were collected from English-language online cancer communities, which may limit the generalizability of the findings to data posted in other languages. Second, perceptions regarding immunotherapy on social media and cancer forums are dynamic, and may be influenced by current news or events. Therefore, future research can consider exploring the fluctuations in posts on these platforms over varying time periods. Our study demonstrated that applying artificial intelligence to analyze social media data can facilitate our understanding of patients' perspectives as well as the effectiveness of machine-learning techniques, specifically BERTopic, in processing substantial volumes of data retrieved from social media platforms and online forums. However, this approach requires validation, although the performance metrics (topic diversity and topic coherence scores) quantified from our models suggest that BERTopic is effective at unveiling latent patterns within large datasets.

5. Conclusions

The social media and health forums is a valuable source of data capturing insights and perspectives of the online communities on cancer treatments. Analyzing social media data through artificial intelligence approaches offers an advantage by bypassing many logistical hurdles associated with traditional approaches and facilitating accelerated and cost-effective data collection and generation. While immunotherapy presents a beacon of hope and a viable treatment option, its side effects and the accompanying potential lifestyle disruptions cannot be ignored. The recurring topics related to the use of immunotherapy identified in this study can serve as valuable tools to enhance communication between clinicians and patients with cancer.

Acknowledgements

XYW and CSL conceptualized the study, conducted the final analysis, wrote the draft, and reviewed the manuscript. MKHH, HHFL, and KRZ conceptualized the study and reviewed the manuscript. CKN conceptualized the study, conducted the analysis and investigation, and reviewed the manuscript. YTC conceptualized the study, conducted the final analysis and investigation, wrote the draft, supervised, and reviewed the manuscript. We would like to thank the Worcester Polytechnic Institute's Turing cluster hosted by Worcester Polytechnic Institute for the implementation of the computing process.

Conflicts of Interest

None declared.

Abbreviations

irAEs: immunotherapy-related adverse events
BERT: Bidirectional Encoder Representation from Transformers

References:

1. Immunotherapy to Treat Cancer. National Cancer Institute. 2019. https://www.cancer.gov/about-cancer/treatment/types/immunotherapy
2. Miller, K. D., Nogueira, L., Devasia, T., et al. Cancer treatment and survivorship statistics, 2022. CA Cancer J Clin, 2022; 72(5): 409-436. https://doi.org/10.3322/caac.21731
3. Tan, S., Li, D., & Zhu, X. Cancer immunotherapy: Pros, cons and beyond. Biomed Pharmacother, 2020; 124: 109821. https://doi.org/10.1016/j.biopha.2020.109821
4. Chen, D. S., & Mellman, I. Elements of cancer immunity and the cancer-immune set point. Nature, 2017; 541(7637): 321-330. https://doi.org/10.1038/nature21349
5. Bien, D. R., Danner, M., Vennedey, V., et al. Patients' Preferences for Outcome, Process and Cost Attributes in Cancer Treatment: A Systematic Review of Discrete Choice Experiments. Patient, 2017; 10(5): 553-565. https://doi.org/10.1007/s40271-017-0235-y
6. Ihrig, A., Richter, J., Bugaj, T. J., et al. Between hope and reality: How oncology physicians and information providers of a cancer information service manage patients' expectations for and experiences with immunotherapies. Patient Education and Counseling, 2023; 109, Article 107622. https://doi.org/10.1016/j.pec.2023.107622
7. Hesse, B. W., Greenberg, A. J., & Rutten, L. J. The role of Internet resources in clinical oncology: promises and challenges. Nat Rev Clin Oncol, 2016; 13(12): 767-776. https://doi.org/10.1038/nrclinonc.2016.78
8. Iftikhar, R., & Abaalkhail, B. Health-Seeking Influence Reflected by Online Health-Related Messages Received on Social Media: Cross-Sectional Survey. J Med Internet Res, 2017; 19(11): e382. https://doi.org/10.2196/jmir.5989
9. Jenei, K., Burgess, M., Peacock, S., et al. Experiences and perspectives of individuals accessing CAR-T cell therapy: A qualitative analysis of online Reddit discussions. J Cancer Policy, 2021; 30: 100303. https://doi.org/10.1016/j.jcpo.2021.100303
10. Booth, A., Manson, S., Halhol, S., et al. Using Health-Related Social Media to Understand the Experiences of Adults With Lung Cancer in the Era of Immuno-Oncology and Targeted Therapies: Observational Study. JMIR Cancer, 2023; 9: e45707. https://doi.org/10.2196/45707
11. Renner, S., Loussikian, P., Foulquie, P., et al. Patient and Caregiver Perceptions of Advanced Bladder Cancer Systemic Treatments: Infodemiology Study Based on Social Media Data. JMIR Cancer, 2023; 9: e45011. https://doi.org/10.2196/45011
12. Chang, A., Xian, X., Liu, M. T., et al. Health Communication through Positive and Solidarity Messages Amid the COVID-19 Pandemic: Automated Content Analysis of Facebook Uses. Int J Environ Res Public Health, 2022; 19(10). https://doi.org/10.3390/ijerph19106159
13. Rodrigues, A., Chauhan, J., Sagkriotis, A., et al. Understanding the lived experience of lung cancer: a European social media listening study [Rodrigues et al.]. BMC Cancer, 2022; 22(1): 475. https://doi.org/10.1186/s12885-022-09505-4
14. Grootendorst, M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv preprint, 2022. https://doi.org/10.48550/arXiv.2203.05794
15. Ng, Q. X., Lee, D. Y. X., Yau, C. E., et al. Public perception on 'healthy ageing' in the past decade: An unsupervised machine learning of 63,809 Twitter posts. Heliyon, 2023; 9(2): e13118. https://doi.org/10.1016/j.heliyon.2023.e13118
16. Ng, Q. X., Yau, C. E., Lim, Y. L., et al. Public sentiment on the global outbreak of monkeypox: an unsupervised machine learning analysis of 352,182 twitter posts. Public Health, 2022; 213: 1-4. https://doi.org/10.1016/j.puhe.2022.09.008
17. Devlin, J., Chang, M. W., Lee, K., et al. Bert: Pre-training of deep bidirectional transformers for language understanding. Proceedings of naacL-HLT, 2019; 1: 2. https://doi.org/10.48550/arXiv.1810.04805

18. Nayak, P. (2019, October 25) *Understanding searches better than ever before*. https://blog.google/products/search/search-language-understanding-bert/
19. Wagner, J. (2021, September 28) *The most popular social networks of 2021*. https://www.ignitesocialmedia.com/twitter-marketing/the-most-popular-social-networks-of-2021/
20. Lam, C. S., Zhou, K., Loong, H. H., et al. The Use of Traditional, Complementary, and Integrative Medicine in Cancer: Data-Mining Study of 1 Million Web-Based Posts From Health Forums and Social Media Platforms. J Med Internet Res, 2023; 25: e45408. https://doi.org/10.2196/45408
21. Kocaman, V., & Talby, D. Spark NLP: natural language understanding at scale. Software Impacts, 2021; 8: 100058. https://doi.org/10.1016/j.simpa.2021.100058
22. Healthcare NLP State of the Art Medical Language Models. John Snow Labs. 2024. https://www.johnsnowlabs.com/healthcare-nlp/
23. Saracco, B. H. Data Science and Predictive Analytics: Biomedical and Health Applications Using R. Journal of the Medical Library Association, 2020; 108(2): 334. https://doi.org/10.5195/jmla.2020.901
24. Grootendorst, M. (2023, September 1) *BERTopic*. https://github.com/MaartenGr/BERTopic
25. Krishna, B. S. V., Satheesh, P., & SuneelKumar, R. Comparative Study of K-means and Bisecting k-means Techniques in Wordnet Based Document Clustering. International Journal of Engineering and Advanced Technology, 2012; 13.2: 127-131. https://doi.org/10.2478/v10038-012-0013-6
26. Dieng, A. B., Ruiz, F. J. R., & Blei, D. M. Topic modeling in embedding spaces. Transactions of the Association for Computational Linguistics, 2020; 8: 439-453. https://doi.org/10.48550/arXiv.1907.04907
27. Mimno, D., Wallach, H., Talley, E., et al. (2011). Optimizing semantic coherence in topic models. *Proceedings of the 2011 conference on empirical methods in natural language processing*. https://aclanthology.org/D11-1024
28. Bonta, V., Kumaresh, N., & Janardhan, N. A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis. Asian Journal of Computer Science and Technology, 2019. https://doi.org/10.51983/ajcst-2019.8.s2.2037
29. Aljedaani, W., Rustam, F., Mkaouer, M. W., et al. Sentiment analysis on Twitter data integrating TextBlob and deep learning models: The case of US airline industry. Knowledge-Based Systems, 2022; 255: 109780. https://doi.org/https://doi.org/10.1016/j.knosys.2022.109780
30. Thapa, D. K., Visentin, D. C., Kornhaber, R., et al. The influence of online health information on health decisions: A systematic review. Patient Educ Couns, 2021; 104(4): 770-784. https://doi.org/10.1016/j.pec.2020.11.016
31. Jacobs, W., Amuta, A. O., & Jeon, K. C. Health information seeking in the digital age: An analysis of health information seeking behavior among US adults. Cogent Social Sciences, 2017; 3(1): 1302785. https://doi.org/10.1080/23311886.2017.1302785
32. Palanica, A., Flaschner, P., Thommandram, A., et al. Physicians' Perceptions of Chatbots in Health Care: Cross-Sectional Web-Based Survey. J Med Internet Res, 2019; 21(4): e12887. https://doi.org/10.2196/12887
33. Suarez-Lledo, V., & Alvarez-Galvez, J. Prevalence of Health Misinformation on Social Media: Systematic Review. J Med Internet Res, 2021; 23(1): e17187. https://doi.org/10.2196/17187
34. Tran, G., & Zafar, S. Y. Financial toxicity and implications for cancer care in the era of molecular and immune therapies. Ann Transl Med, 2018; 6(9): 166. https://doi.org/10.21037/atm.2018.03.28
35. Kovacs, D., Kovacs, P., Eszlari, N., et al. Psychological side effects of immune therapies: symptoms and pathomechanism. Curr Opin Pharmacol, 2016; 29: 97-103. https://doi.org/10.1016/j.coph.2016.06.008
36. Lacouture, M., & Sibaud, V. Toxic Side Effects of Targeted Therapies and Immunotherapies Affecting the Skin, Oral Mucosa, Hair, and Nails. Am J Clin Dermatol, 2018; 19(Suppl 1): 31-39. https://doi.org/10.1007/s40257-018-0384-3
37. Haanen, J., Obeid, M., Spain, L., et al. Management of toxicities from immunotherapy: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. Ann Oncol, 2022; 33(12): 1217-1238. https://doi.org/10.1016/j.annonc.2022.10.001

38. Brahmer, J. R., Lacchetti, C., Schneider, B. J., et al. Management of Immune-Related Adverse Events in Patients Treated With Immune Checkpoint Inhibitor Therapy: American Society of Clinical Oncology Clinical Practice Guideline. J Clin Oncol, 2018; 36(17): 1714-1768. https://doi.org/10.1200/JCO.2017.77.6385
39. Rahman, M. M., Behl, T., Islam, M. R., et al. Emerging Management Approach for the Adverse Events of Immunotherapy of Cancer. Molecules, 2022; 27(12). https://doi.org/10.3390/molecules27123798
40. Zhang, N., & Xiao, X. H. Integrative medicine in the era of cancer immunotherapy: Challenges and opportunities. J Integr Med, 2021; 19(4): 291-294. https://doi.org/10.1016/j.joim.2021.03.005
41. Jia, W., & Wang, L. Using Traditional Chinese Medicine to Treat Hepatocellular Carcinoma by Targeting Tumor Immunity. Evid Based Complement Alternat Med, 2020; 2020: 9843486. https://doi.org/10.1155/2020/9843486
42. Schneider, B. J., Naidoo, J., Santomasso, B. D., et al. Management of Immune-Related Adverse Events in Patients Treated With Immune Checkpoint Inhibitor Therapy: ASCO Guideline Update. J Clin Oncol, 2021; 39(36): 4073-4126. https://doi.org/10.1200/JCO.21.01440

Preprint JMIR Publications

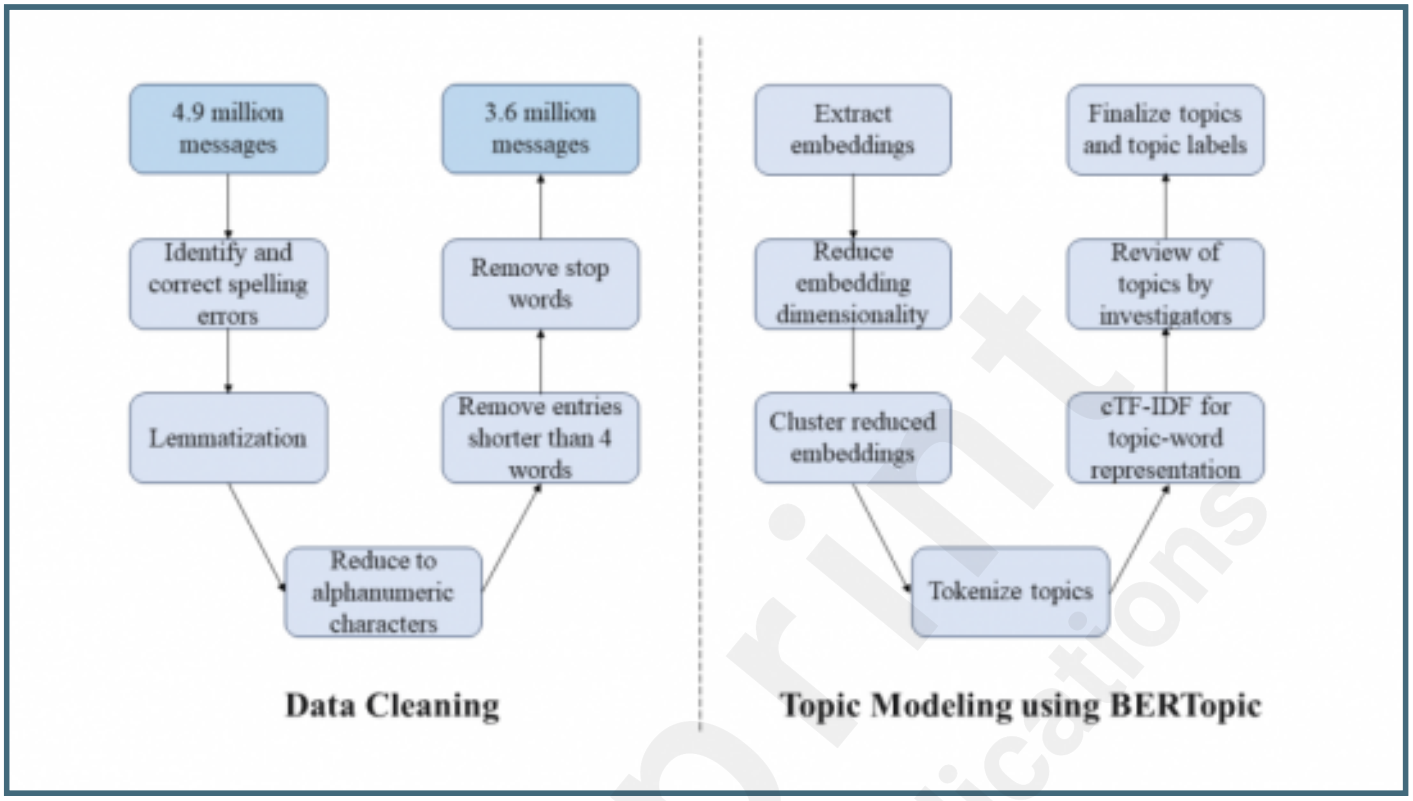
Supplementary Files

Preprint JMIR Publications

Figures

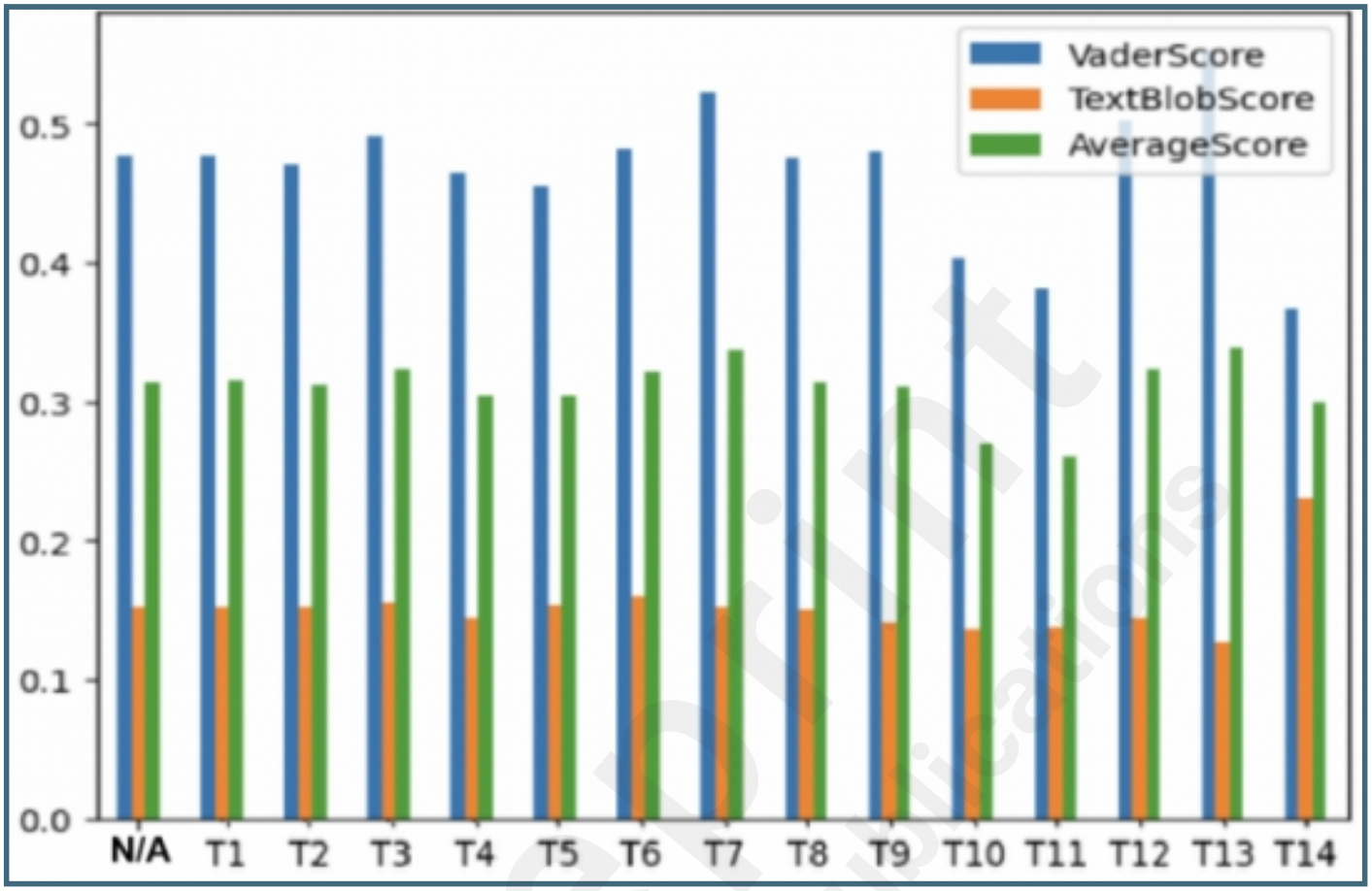
Preprint JMIR Publications

Flowchart depicting the data mining process.



Preprint JMIR Publications

Sentiment analysis results.



Preprint JMIR Publications

Multimedia Appendixes

Preprint JMIR Publications

List of online health forums selected in the data extraction.
URL: http://asset.jmir.pub/assets/40a2496f748542f4cf32d115fef3d1cb.doc

Example of keywords related to immunotherapy for post/thread search and the data sources.
URL: http://asset.jmir.pub/assets/8ac812d60247bffe863788ee6673e8a4.doc

The performances of BERTopic on different numbers of topics.
URL: http://asset.jmir.pub/assets/0a3b0cf534de663b63887c94a8f59ff1.doc

The distribution of topic quality scores.
URL: http://asset.jmir.pub/assets/280317b1840d2f575ab69bd0f6e82a2c.png

Preprint JMIR Publications