ORIGINAL RESEARCH

What are IBD Patients Talking About on Twitter? Using Natural Language Understanding to Investigate Patients' Tweets

Maya Stemmer[1] · Yisrael Parmet[1] · Gilad Ravid[1]

Abstract

This research aims to investigate what patients with inflammatory bowel disease (IBD) are talking about on Twitter and learn from the experimental knowledge they share online. The study presents a framework for analyzing patients' tweets and comparing their content to tweets of the general population. We started by constructing two datasets of tweets—a dataset of patients' tweets and a control dataset for comparison. Then, we thematically classified the tweets and obtained a subset of tweets related to health and nutrition. We used a Dirichlet regression to compare the thematic segmentations of the two groups. We continued by extracting keywords from the filtered tweets and applying entity sentiment analysis to determine the patients' sentiments towards the extracted keywords. Finally, we detected emotions within the tweets and used a Wilcoxon test to compare the emotions conveyed in each group. We found statistically significant differences between the patients' thematic segmentations and those of the control group and observed significant differences in the emotions each group expressed while talking about health. Not only do patients talk more about health in comparison to the general Twitter population, but they also address the subject with negative sentiments and express more negative emotions. The personal information IBD patients share on Twitter can be used to derive complementary knowledge about the disease and provide an additional foundation to existing medical research on IBD. The four stages of the study are also feasible to extend to other chronic conditions.

Introduction

Background

In recent years, social networking sites and online communities have served as alternate information sources for patients. Patients everywhere use social media to share health and treatment information, learn from each other's experiences, and provide social support. Particularly, chronically ill patients, who use online environments to support daily self-management, exchange experiential knowledge about their disease that extends far beyond medical care. Mining these informative conversations may shed some light on patients' ways of life and support the research of chronic conditions.

Inflammatory bowel disease (IBD) is a chronic inflammation condition of the digestive system characterized by flares and remission states. The two primary diseases identified with IBD, Crohn's disease (CD) and ulcerative colitis (UC), are usually diagnosed in young patients (in the age range of 15–30 years). The incidences of IBD are rapidly increasing, and it has evolved into a global disease [1–4].

There are no medications or surgical procedures that can cure IBD. Treatment options can only help with symptoms affecting each patient differently. They involve prescription drugs and lifestyle change solutions, such as diets and therapies. Symptoms include abdominal pain, diarrhea, and fatigue; severe cases may result in hospitalization or surgical interventions [5, 6].

As chronic bowel diseases, both CD and UC require day-to-day care for drug consumption and special nutrition. Patients must adhere to strict dietary regimes and maintain

✉ Maya Stemmer
   mayast@post.bgu.ac.il

1   Ben-Gurion University of the Negev, P.O.B. 653, 8410501 Beer-Sheva, Israel

a calm routine. Changes in nutrition or physical activity, currently tested by trial and error, result in a long and excruciating process for the patients [7–9]. By collecting and analyzing patients' data on social media, we can learn from their personal experiences and support existing medical knowledge regarding the disease.

Twitter is a powerful tool for disseminating health-related information and an accessible platform for patients needing immediate social support or relief [10]. By overcoming space and distance, patients on Twitter form a community that disregards physical boundaries or immobility. A sense of common ground can help break down barriers and enable conversation, increasing a person's willingness to share [11, 12]. It may be easier to consult other patients who can relate and better understand the situation based on personal experience. One can identify more closely with users' stories similar to one's own and embrace their advice more easily [13]. When people disclose health information on Twitter, they expose themselves to various opinions and reduce uncertainty about their disease [14].

Patients describe IBD as an embarrassing disease that causes immediate disruption of daily activities. They experience difficulties adjusting to the forced changes and consider themselves different from their peers. Since IBD is identified with frequent bowel movements, people do not hasten to share their disease with others [7, 9]. According to IBD patients, part of the embarrassment can be attributed to a lack of public awareness. Outsiders cannot see that a person's stomach hurts or his bowels are scarred. The disease is invisible, and others might doubt it exists [15, 16].

The embarrassment caused by IBD and the need to confide in people who undergo similar experiences help explain the creation of IBD-related communities on Twitter. Patients with IBD are the most common type of users that engage in IBD-related discussions on Twitter [17, 18]. They use Twitter to share personal experiences and to seek social support. They exchange thoughts about symptoms and medications and recommend treatments to one another [19]. By sharing their experience with the disease on Twitter, patients fight disease invisibility and raise public awareness about IBD [20].

Objective and Contribution

Our research aimed to investigate what patients with IBD are talking about on Twitter and to learn from the experimental knowledge of living with the disease they share online. We wished to identify lifestyle-related treatments IBD patients endure to maintain their disease and to determine their sentiments towards them. We collected and analyzed tweets by IBD patients to identify lifestyle-related keywords and their influence on the patients.

This manuscript details our methodology in four main stages to improve clarity and reproducibility. First, we describe how we collected and prepared our data. We obtained a primary dataset containing patients' tweets and a control dataset for comparison. Then, we classify the tweets into categories to focus our research on tweets related to health and nutrition. We continue by extracting keywords from the filtered tweets and determining the patients' sentiments toward them. Finally, we detect emotions within the subset of tweets and reflect on patients' moods. Within each stage, we conduct statistical analyses to compare the patients and the control group.

In a previous study [21], we identified 337 IBD patients who openly tweeted about their disease on Twitter. In this study, we used Twitter API [22] to collect their tweets and analyze the content they shared about IBD. Using the Natural Language Understanding (NLU) module by IBM Cloud [23], we applied category classification to the tweets and identified themes discussed by IBD patients on Twitter. To account for the general Twitter population, we compiled a control group of random users and compared their thematic segmentation with one of the IBD patients. Our analysis showed statistically significant differences between the thematic segmentations of the two groups: while IBD patients talked more about health, fitness, and nutrition in comparison to the general population, the latter increased to refer to politics and society.

Based on the thematic segmentation, we derived a collection of patients' tweets related to health or nutrition and investigated the index terms they contained. We used the keyword extraction feature of the NLU module to extract keywords from the tweets and calculate their sentiments within the text. We performed several text-cleaning procedures to refine the extracted keywords and obtained the aggregated sentiment of each keyword. Visualizing the keywords and their mean sentiment with word clouds showed that patients' sentiments turned more negative when they talked about their disease.

We further investigated tweets related to health or nutrition and used the NLU module to detect five emotions within the text: sadness, joy, fear, disgust, and anger. They are considered "universal" emotions that successfully capture the different types of human feelings [24]. Our statistical analysis showed that patients tend to express more sadness and fear when discussing health compared to the control group.

To adhere to ethical norms and maintain user privacy, we only publish aggregated results that do not reveal the specific users. The examples containing direct quotes from tweets are presented here after obtaining informed consent from their authors.

The study suggests that the personal information shared by IBD patients on Twitter can be used to understand better the disease and how it affects patients' lives. By applying

text mining and statistics to patients' tweets, we can learn about the different treatments they try and identify helpful treatments. For example, one might expect that sentiment towards a relaxing activity like yoga will be positive, while sentiment towards alcohol, which can irritate the stomach, will be negative. Therefore, findings from such research can provide complementary knowledge about the disease based on the wisdom of the crowd. The methods can be adapted to other diseases and enhance medical research on chronic conditions.

The rest of this paper is organized as follows: in "Related work", we explore related research motivating our study; in "Methods", we describe in detail the methods used in this study; in "Results", we review its results; in "Discussion", we discuss the implications of the results and suggest future research; and in "Conclusions", we summarize and conclude.

Related Work

Recently, substantial research has been dedicated to thematically analyzing health-related tweets. Researchers conducted large-scale analyses of tweets related to diabetes [25], COVID-19 [26, 27], cancer [28], and IBD [29, 30]. In contrast to this study, they did not focus their research on patients' tweets. Rather, they collected relevant tweets via Twitter APIs and investigated public opinion. In other studies exploring relatively small-scale datasets of tweets, the researchers were able to identify patients [31–33]. However, the patient identification process and the thematic segmentation of the tweets were done manually. In this study, we used a predefined set of patients and classified the tweets into categories using an automated tool. Furthermore, all previous studies obtained their datasets by searching Twitter for relevant keywords or hashtags. In contrast, we performed thematic analyses of patients' full timelines.

Unlike previous Twitter research [34, 35], we did not focus our work on patients' discussion within their disease domain or compare patients to other involved parties. We investigated patients' general discourse on Twitter and compared it with the public discourse. We suggested a method for sampling the general Twitter population and forming a control group to which we compared patients' results. Our control group was composed of random Twitter users unnecessarily related to the disease. The innovative method allowed us to perform a thematic segmentation of all patients' tweets and highlight the fraction they devoted to health and nutrition.

The concept of forming a control group containing random users was used in two previous studies monitoring depression on Twitter [36, 37]. They used Twitter APIs

to collect a group of users who self-reported suffering from depression and a control group representing the general public. To form the control group, Chen et al. [37] streamed one day of tweets containing the keyword "the." They collected their authors after double-checking that none of them overlapped with the depression group. Zhang et al. [36] randomly selected users who did not use depression-related terms in their past 200 tweets or descriptions. Their control group contained the exact number of users as their depression group. Like their studies, we limited the search to tweets written in English, excluded retweets, and filtered out organizations and spammers. However, we streamed tweets over different periods to account for different daytimes or working hours worldwide.

Previous research on emotion detection on Twitter focused on building the emotion extraction framework [38, 39]. In this study, we utilized a pre-trained emotion classifier and applied emotion detection to Twitter data in the domain of health and fitness. We demonstrated how patients' feelings differ from those of the general population as they are affected by their disease.

The emotions expressed in each group (patients vs. control) were compared without considering changes over time. We averaged the intensity of each emotion by a user and then compared the groups (patients vs. control) using a two-independent-sample Wilcoxon (Mann–Whitney) test. The researchers in [36] obtained a single emotion score for each user by consolidating their entire set of tweets into one document. Then, they illustrated the emotion distribution of each group. In [37], the authors suggested a method for comparing emotions between the main and control groups over time which we did not apply here.

The most similar work to ours was carried out by Perez et al. [40], who conducted a large-scale analysis of IBD-related tweets. They started with a dataset of IBD-related tweets in which they automatically identified the ones written by patients. They further investigated the subset of patients' tweets and automatically classified the tweets' topics and sentiments. Other than the two differences mentioned above (i.e., dataset collection and discussion comparison), their study differs from ours in another way, as they investigated the sentiment of the entire tweet.

This study extracted keywords from patients' tweets within different thematic categories and calculated their averaged sentiments. We applied entity sentiment analysis and investigated patients' sentiments toward lifestyle-related keywords. In a previous study [41], we investigated the sentiments towards keywords from a predefined list containing only single words. Here, we extended the analysis to index terms of both words and phrases extracted directly from the text. To compare the sentiments between the categories, we used independent samples t-tests. This method was also applied by Gabarron et al. [42].

Methods

Overview

This study presents a pipeline for analyzing patients' tweets and comparing their content with tweets of the general Twitter population. The study was conducted in four main stages, as detailed in the following sections. In this introductory section, we give an overview of the main stages and illustrate the steps of the study in Figs. 1 and 2. "Data collection and preparation" explains how we constructed two datasets of tweets—the main dataset containing patients' tweets and a control dataset containing tweets from random Twitter users. The two datasets provided the foundation for our analyses in "Category Classification", "Keyword Extraction", and "Emotion Classification". Figure 1 illustrates the data collection process.

As described in "Category Classification", we started by classifying the tweets into categories and obtaining the

thematic segmentations of the tweets from both datasets. We compared the results between the patients and the control group using a Dirichlet regression. Then, we gathered tweets related to health and nutrition for further investigation. In "Keyword Extraction", we explored patients' tweets about health and nutrition, identified frequently used keywords, and obtained their sentiments toward them. In "Emotion Classification", we detected emotions within the tweets and compared the patients' results with those of the control group using a Wilcoxon test. Figure 2 illustrates these three steps of the data analysis process.

Data Collection and Preparation

Patients' Group

In a previous study [21], aiming to identify patients with IBD on Twitter, we trained and evaluated a classifier that distinguishes patients with IBD from other users who tweet about the disease. Our manual labeling process revealed 337
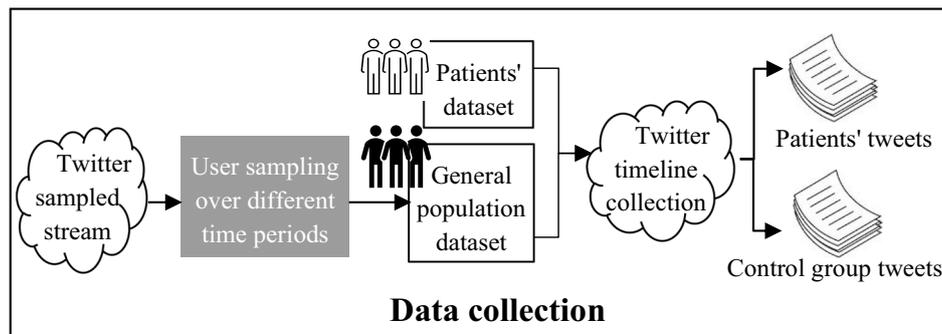

Fig. 1 An illustration of the first stage of the study—collecting two datasets of tweets, a main dataset containing patients' tweets and a control dataset containing tweets of random Twitter users. The patients' group was obtained in an earlier study. The control group was obtained by sampling Twitter Stream API over different time periods and randomly collecting private accounts. Then, we used Twitter Search API to collect the timelines of the users and obtained two datasets of tweets

Fig. 2 An illustration of the second, third, and fourth stages of the study—classifying the tweets into categories and focusing on tweets related to health and nutrition; extracting keywords from patients' tweets and determining the patients' sentiment toward them; and detecting emotions within the tweets. After each stage, we perform statistical analyses to compare between the patients' group and the control group and learn on patients' behavior
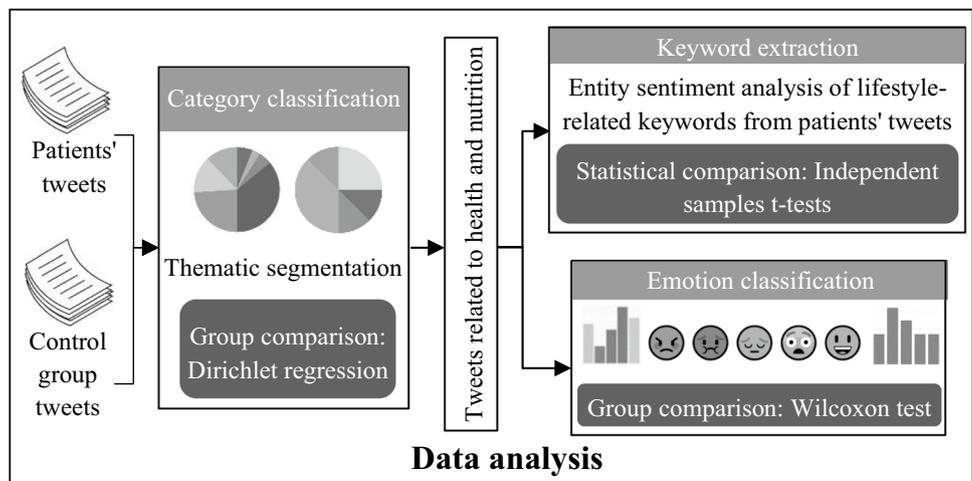
IBD patients who publicly declared their disease on Twitter in a tweet written in English. These 337 patients formed the dataset for our current study.

On November 12th, 2020, we used Twitter Search API to collect the Twitter timelines of the 337 patients we identified. The Search API allows one to collect up to 3,200 of a user's latest tweets. If the user has less than 3,200 tweets, one may collect their complete timeline using the API. However, if the user tweets frequently and has more than 3,200 tweets, one may only collect their latest. We wanted to investigate the original content shared by patients, so we excluded retweets (RTs) from the search. The process resulted in a collection of 628,301 tweets by IBD patients, sporadically written since the creation of Twitter in 2006.

To further analyze the tweets, we cleaned their text by removing all screen names (identified by the @ character) and URLs and turning all text lowercase. We chose not to omit emojis, special characters, or punctuation since they were valuable for sentiment analysis. Examples of the cleaning process can be found in Table 1.

The purposes of the analysis were to understand what IBD patients are talking about on Twitter, identify the lifestyle-related treatments that help manage their disease, and determine the patients' sentiments towards them. We used the Natural Language Understanding (NLU) module by IBM Cloud [23] to apply category classification, keyword extraction, and emotion classification to the clean tweets. After text cleaning, 78,140 (12.4%) tweets contained less than three words and were too short for language processing. The remaining dataset of 550,161 longer tweets was given as input to the NLU module for category classification.

Control Group

To evaluate the results by comparing the results of the patients with those of the general Twitter population, we wished to form a control group of Twitter users. Since demographic data are unavailable on Twitter, we used Twitter's Sampled Stream API to collect a group of random users to account for the general population.

Twitter's Sampled Stream API streams about 1% of all tweets in real time [43]. To account for different daytimes and working hours over the world, we used the API to stream users who wrote new content in English (and not just retweeted) over 3 days (Sunday, May 23rd, 2021; Tuesday, May 25th, 2021; and Thursday, May 27th, 2021) and 3-day periods (7 AM, 12 PM, 8 PM UTC). We filtered out

Table 1 Three examples of text cleaning and category classification. For each original tweet, we present its cleaned text (middle column) and the categories and subcategories it was classified into by the NLU module (right column)

	Original text	Text after cleaning	Text category, likelihood
1	As soon as I can eat reliably and have energy I'm making a couple cheesecakes	As soon as i can eat reliably and have energy i'm making a couple cheesecakes	**/Food and drink**, 0.717184
			/Society/social institution/divorce, 0.671966
			/Food and drink/food, 0.650148
			/Society/social institution/marriage, 0.618626
			/Pets, 0.578122
			/Family and parenting, 0.549967
			/Family and parenting/children, 0.537206
			/Health and fitness/weight loss, 0.533751
			/Food and drink/vegetarian, 0.530938
			/food and drink/cuisines, 0.527833
2	@bottomline_ibd great poll. I do have the odd binge, but IBD has changed what I can drink. No more red wine or ale 😔	Great poll. i do have the odd binge, but ibd has changed what i can drink. no more red wine or ale 😔	**/Food and drink** /beverages /alcoholic beverages /cocktails and beer, 0.84566
			/Health and fitness /disease /ibs and crohn's disease, 0.80538
			/Food and drink /beverages /alcoholic beverages /wine, 0.80134
			/Food and drink /beverages /non alcoholic beverages /soft drinks, 0.54913
			/Food and drink /food allergies, 0.53412
3	I am living proof that yoga can help #uchicagoibd #studiothree #yoga #ibd	I am living proof that yoga can help #uchicagoibd #studiothree #yoga #ibd	**/Society**/senior living, 0.922317
			/Religion and spirituality /hinduism, 0.902807

The categories are sorted by decreasing likelihood, and the most likely high-level category is marked with bold. Notice that the original tweet in the second example starts with a screen name, as indicated by the @ sign. Therefore, the screen name was removed during the text's cleaning process

organizations and spammers and reached a control group of 57 private accounts of random English speakers worldwide.

We repeated the process we applied for the patients and performed a similar category analysis for the control group: we collected their timelines (97,439 tweets overall), cleaned all the tweets, removed those with less than three words after cleaning, and used the NLU module to classify the 78,665 remaining tweets into categories.

Category Classification

Patients and Control Group

The category classification feature of the NLU module aims to identify the text's theme. Given a text, the NLU module lists possible categories and subcategories and their corresponding likelihood. The likelihood score represents the confidence of the classification, and only those with a confidence level higher than 0.5 are returned. The module uses a predefined list of 23 high-level categories with a different number of subcategories in each group, to which it classifies the text, if applicable. The returned list of possible classifications is sorted based on their likelihood, from largest to smallest (see examples in Table 1).

We used the module to apply category classification to each tweet in our dataset. For the sake of our analysis, we only addressed the module's first and most likely classification and returned all classifications to their high-level categories (marked with bold background in Table 1). We repeated the process for all the tweets in our dataset and aggregated the results. We obtained a segmentation of the themes discussed by patients over the 23 categories and a similar segmentation for the control group. In each group, we derived a segmentation for all the users and each separately.

Comparing the Groups

After obtaining a segmentation of the categories discussed in each group (patients vs. control), we used a Dirichlet regression [44] to compare the segmentations of the two groups. The Dirichlet regression approach was suitable for the compositional data we gathered [45]. Each user was represented by a vector of their segmentation over the 23 categories, and we wished to compare the vectors between the groups (and not just one value). Therefore, the data we gathered per user were compositional data, where each instance is composed of a set of positive variables that sum up to 1. A Dirichlet regression was suitable here since the Dirichlet distribution is a continuous multinomial distribution that allows for nominal categories without the limitation of choosing a single category each time.

For each user, we obtained their category segmentation and annotation of their group. Using the 'DirichletReg' package in R [46], we fitted a Dirichlet regression model to the data. The dependent variable was the segmentation of each user over the 23 categories, and the independent variable was the group membership—whether the user belonged to the patients' group or to the control one.

Aiming to test whether the group membership influenced the category segmentation, we trained the model twice and compared two statistical models: the full model (with the group indicator as an explanatory variable) and the null model (without the group indicator). In the full model, each of the 23 categories was represented by an intercept and the group variable, resulting in 46 parameters. In the null model, each category was represented only by an intercept. Therefore, the null model contained 23 parameters overall. Finally, we used three well-known statistical measures to compare the results of the two models: a likelihood-ratio test (LRT) [47], Akaike information criteria (AIC), and Bayesian information criteria (BIC) [48, 49].

Keyword Extraction

The next goal was to identify the lifestyle-related treatments patients endure to manage their disease and to determine their sentiments towards them. Since we were interested in keywords that reflected patients' lifestyles, we utilized the obtained segmentation from the "Category Classification" stage to focus our analysis on keywords related to health and nutrition. We grouped together all tweets that were categorized by the NLU module as related to "health and fitness" (53,598 tweets) or "food and drink" (30,747 tweets) and added all other tweets that explicitly mentioned IBD by containing at least one of the keywords: *crohn*, *colitis* or *#IBD* (additional 1228 tweets). Overall, 85,573 tweets were selected for further analysis of keyword extraction.

The keyword extraction feature of the NLU module recognizes words and phrases of high importance within the text and calculates their sentiments. Given a text, the NLU module returns a list of keywords sorted by diminishing relevance to the text and their corresponding sentiment. For each keyword, the module provides a score on the closed line segment $[-1,1]$ representing the keyword's sentiment within the text. -1 represents highly negative sentiment, and 1 represents exceptionally positive sentiment; a score of 0 means that the keyword was mentioned in a neutral context.

As can be seen in the examples in Table 2, the extracted keywords can be very specific and may include adjectives. We wished to group different variations of similar keywords together and obtain their aggregated sentiment. Therefore, we performed several steps to refine the keywords and achieve this goal:

Table 2 Two examples of keyword extraction and refinement

	Text after cleaning	Extracted keyword, sentiment	Fixed keyword, sentiment
1	As soon as i can eat reliably and have energy i'm making a couple of cheesecakes	Couple of cheesecakes, 0.86 energy, 0.86	Cheesecake, 0.86 energy, 0.86
2	Great poll. i do have the odd binge, but ibd has changed what i can drink. no more red wine or ale 😊	Odd binge, − 0.45 Red wine, − 0.83 Great poll, 0.96 Ibd, − 0.45 Ale, − 0.83	Binge, − 0.45 Red wine, − 0.83 Poll, 0.96 Ibd, − 0.45 Ale, − 0.83

For each cleaned tweet, we present the keywords originally extracted by the NLU module and their corresponding sentiments (middle column) and the final versions of the keywords after applying our refinement process (right column)

1. Articles and pronouns: we omitted definite and indefinite articles, demonstrative adjectives and pronouns, and possessive pronouns from all keywords.
2. Quantity: we erased the beginning of keywords containing the combination "_____ of _____" to reflect on quantity (e.g., "300 g of almonds" or "3 cups of coffee"). We kept the part following the "of" (resulting in almonds and coffee, respectively).
3. Adjectives: we wished to include phrases such as "red wine" or "sweet potato," for which the adjectives define their meanings, but to avoid terms such as "amazing cake" or "yummy chocolate," for which the adjectives should be considered in sentiment calculation only. Notice how the first word in all four examples is an adjective. Therefore, Part of Speech tagging was not suitable for performing the differentiation between the phrases. Instead, we combined lists of adjectives from several sources and created a designated adjective list suited for our unique task. Based on the list, we eliminated adjectives that reflected the quality or quantity of the objects (such as awful, big, and yummy). We kept the ones that altered the meaning of the objects and were, therefore, crucial for our analysis (such as hot, spicy, red, and green).
4. Lemmatization: we used spaCy library for NLP in python to apply lemmatization to the text and return the keywords to their basic form.

After completing the keywords refinement, we grouped identical keywords together and calculated the average sentiment for each. We wanted to avoid rare keywords that hardly appeared in our dataset, so we included only keywords that were mentioned at least ten times. To prevent a single patient from sealing a keyword's fate by praising it or denigrating it, we only included keywords mentioned by at least five patients.

We examined the frequently mentioned keywords used by patients in four different groups: keywords from all tweets, keywords from tweets related to health and nutrition, keywords from tweets related to food and drink, and keywords from tweets associated with IBD. There was an overlap between the "health and fitness" category and the tweets related to IBD, as 88.8% of the tweets explicitly mentioned IBD were classified by the NLU module as related to health and fitness. For each group, we generated a word cloud that visualized the frequent keywords within the group. We obtained the 25 most frequent keywords in each group and compared the results using six independent samples t-tests, one for every two groups.

Emotion Classification

Patients and Control Group

The emotion classification feature of the NLU module detects five emotions within a given text: anger, disgust, fear, joy, and sadness. For each emotion, the NLU module provides a score on the closed segment [0,1], representing the degree to which the emotion is conveyed in the text (see examples in Table 3).

We used the module to apply emotion classification to all the tweets in our dataset that were classified as "Health and fitness" or "food and drink" in the "Category Classification" stage. Overall, the module was applied to 53,598 patients' tweets related to "health and fitness" and 30,747 patients' tweets related to "food and drink." In the control group, the emotion classification feature was applied to 3537 tweets related to "health and fitness" and 2870 tweets related to "food and drink."

We aggregated the results per user and obtained the average intensity they conveyed within all their tweets for each emotion. We also aggregated the results per category to assess the overall differences in emotions between the patients and the control group.

Comparing the Groups

In contrast to the category classification conducted earlier, we could not fit a Dirichlet regression to these data. The

Table 3 Two examples of emotion classification

	Text after cleaning	Emotion: intensity
1	As soon as i can eat reliably and have energy i'm making a couple of cheesecakes	Sadness: 0.5423 Joy: 0.2663 Fear: 0.0505 Disgust: 0.0192 Anger: 0.0279
2	Great poll. i do have the odd binge, but ibd has changed what i can drink. no more red wine or ale 😔	Sadness: 0.3679 Joy: 0.6100 Fear: 0.0319 Disgust: 0.0215 Anger: 0.0110

For each cleaned tweet, we present the emotions extracted by the NLU module and their corresponding intensity. The module explores five emotions: sadness, joy, fear, disgust, and anger. Their intensity represents the degree to which it was conveyed in the text

scores of the five emotions were not compositional data as they did not sum up to 1. Instead, each emotion had its own standalone scale, from 0 to 100 percent, and we had to compare each separately.

We used a two-independent-sample Wilcoxon (Mann–Whitney) test [50] to compare the results between the two groups—patients vs. control, and the two categories—"health and fitness" and "food and drink." For each user, we obtained the averaged intensities of emotions conveyed in each category and annotation of the group they belonged to (patients or control). Using the wilcox.test function in R, we compared the averaged intensity between the groups for each emotion and each category.

Results

Category Classification

Patient's Group

The NLU module was used for category classification in 550,161 tweets written by IBD patients. 7964 of the tweets were ignored by the NLU module since they did not contain enough data for category classification. Another 4830 tweets were classified as written in languages other than English. A category was successfully retrieved for 537,367 tweets (97.7%).

15.8% of the categorized tweets were classified as "art and entertainment," the most frequent category. 11.9% were classified as "society", 10.0% as "health and fitness", and 9.6% were classified as "sports". The classification continued with "law, govt, and politics," accounting for 7.9% of the tweets. It was followed by "family and parenting" with 5.7%, "food and drink" with 5.7%, and "technology and computing" with 5.0% of the tweets. The next four categories: "style and fashion", "automotive and vehicles", "business and industrial", and "pets" accounted for 4.0, 3.3, 3.2 and 3.0%, respectively. The remaining tweets were classified into 11 small categories (of size ranging between 0.2 and 2.5%), accounting for 15% of the tweets. All the 23 high-level categories suggested by IBM were detected within the patients' tweets. The complete category segmentation can be found at https://www.bgu.ac.il/~mayast/IHAW_Appendix.pdf.

In the context of lifestyle, it is interesting to examine three categories: "health and fitness," which captures tweets related to diseases or physical activity; "food and drink," which captures tweets related to nutrition; and "religion and spirituality," which captures tweets related to spiritual, relaxing activities such as yoga and meditation. The "health and fitness" category was the third largest category discussed by patients. The segmentation showed that they talk about these subjects 10% of the time. Another dominant category was "food and drink," holding 5.7% of all tweets as the 7th largest category. Surprisingly, the "religion and spirituality" category was one of the smaller categories, with only 2.5% of the tweets.

Control Group

The NLU module was used for category classification in 78,665 tweets written by users from the control group. After eliminating tweets that did not contain enough data or those classified as written in languages other than English, a category was successfully retrieved for 73,817 tweets (93.8%).

As in the patients' group, the two largest categories of the control group were "art and entertainment" and "society," which accounted for 16.3 and 14.7% of the tweets, respectively. Even though the classification stayed the same, both categories increased in the control group's segmentation.

Continuing to the third largest category, we noticed a difference between the groups in the classification itself. Holding 10% of the patients' tweets, "health and fitness" was the third largest category within the patients' group. With only 4.8% of the random users' tweets, the same category dropped to 6th in the control group's segmentation.

The third largest category within the control group was "law, govt and politics," which held 10.8% of the tweets.

The "food and drink" category dropped three places compared to the patients' group: from 7th place with 5.7% of the patients' tweets to 10th place with only 3.9% of the random users' tweets. The findings show that patients talk more about health and fitness than the control group, who talk more about politics and society. The complete category classification of the control group's tweets can be found at https://www.bgu.ac.il/~mayast/IHAW_Appendix.pdf.

Comparing the Groups

Observing the differences between the categories discussed by patients and those discussed by random users, we used a Dirichlet regression to check whether these differences were statistically significant. We obtained a category segmentation per user and tested whether the user group (patients or control) was a statistically significant explanatory variable.

Table 4 Analysis of deviance table between the two regression models: model 1 (the full model) with the group indicator as an explanatory variable and model 2 (the null model) without

Model	Deviance	N. parameters	Difference (LRT)	df	Pr(> Chi) (p value)
1—full	− 46,766	46			
2—null	− 46,546	23	219.83	23	< 2.2e−16***

The complete model provides smaller deviance, and the results are highly significant ($P < 0.001$)

ns $p > 0.05$; *$p \leq 0.05$; **$p \leq 0.01$; ***$p \leq 0.001$

The group indicator was found to be statistically significant by three different statistical measures: the likelihood-ratio test ($\chi^2(23) = 218.93$) comparing the two statistical models showed $P < 0.001$, and both AIC and BIC were smaller for the model with the group variable. Table 4 presents the analysis of deviance between the two statistical models we compared: model 1 (the full model), containing the user group as an explanatory variable and an intercept, and model 2 (the null model), containing an intercept only.

Figures 3 and 4 visualize the results of the Dirichlet regression we performed. Figure 3 shows the predicted category segmentation for each group: the patients' predicted probabilities are represented by green circles and those of the control group by red diamonds. We can already notice the difference between the two groups regarding health and fitness, as the predicted probability for this category within the patients' group is higher than the equivalent within the control one. On the other hand, we can see that the control group talks more about politics and society than the patients' group. Figure 4 shows the scaled difference between the predicted values of the patients' group and the predicted values of the control group in each category. The scaled difference was computed as the ratio of the difference between the patients and the control to their average. Again, we can see how patients talk more about health and fitness or even food and drink and less about politics and society in comparison to the control group.

Keyword Extraction

Out of the 85,573 patients' tweets selected for keyword extraction, the NLU module identified keywords in 82,778

Fig. 3 Predicted probability of each category by group: the patients' group is marked with green circles and the control group with red diamonds. The patients talked more about health and fitness and food and drink. In contrast, the control group talked more about society and politics
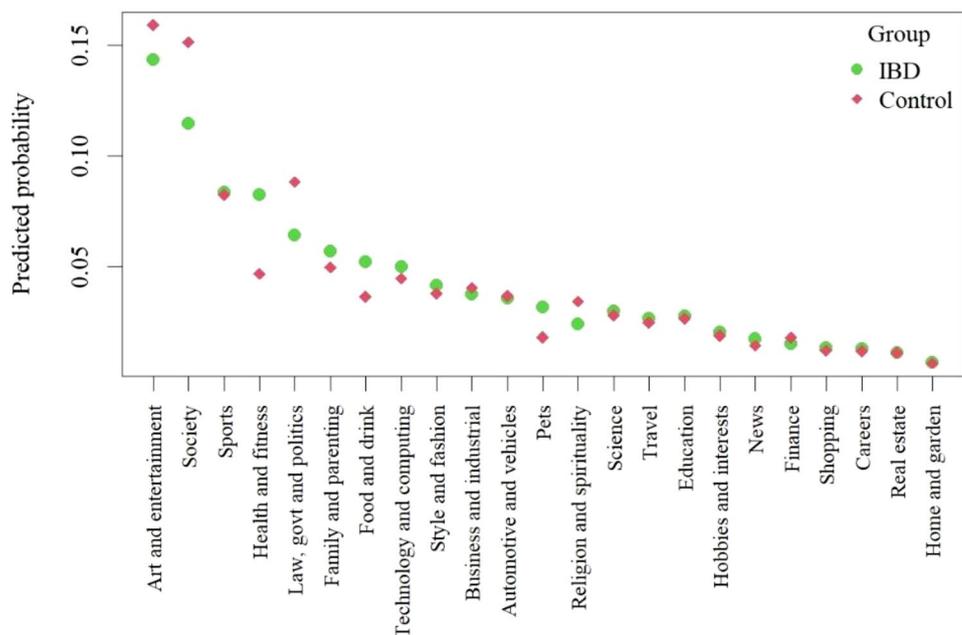
Fig. 4 Scaled differences between the predicted values of the patients' group and the predicted values of the control in each category. Green lines, indicating positive differences, signal that the patients talked more about the respective subjects. Red lines, indicating negative differences, signal that the patients talked less about the respective topics. Substantial differences can be identified in favor of the patients in the "health and fitness" and "food and drink" categories
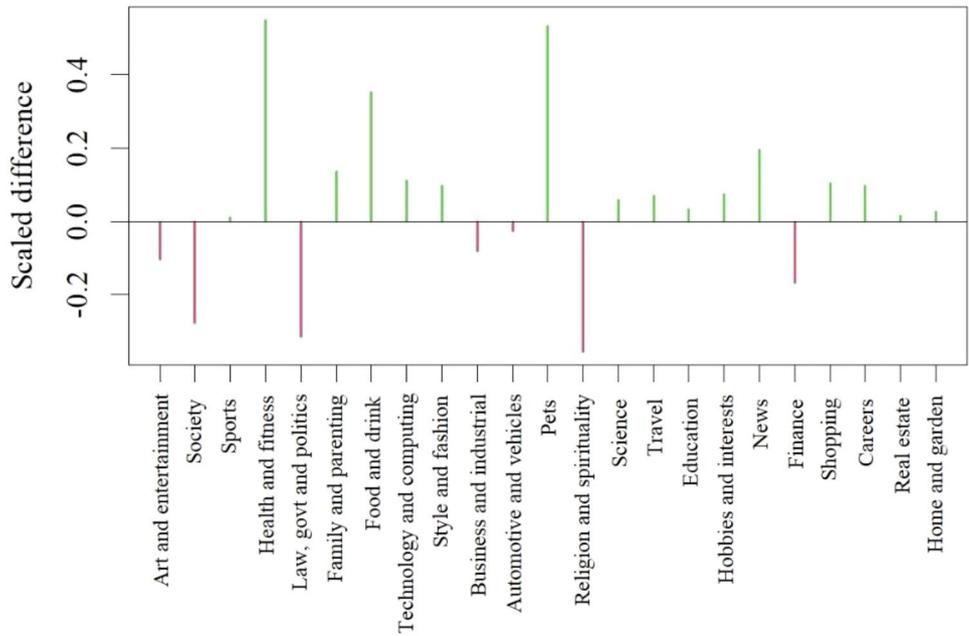
tweets (96.7%), and a total of 218,592 keywords were extracted. After completing the refinement process and aggregating similar keywords, 64,020 different keywords were found. However, only 2553 keywords were mentioned at least ten times in our database, and only 2335 keywords were also mentioned by at least five different patients.

Figure 5 shows word clouds of frequently mentioned keywords generated over four different groups: keywords from all tweets, keywords from tweets related to health and nutrition, keywords from tweets related to food and drink, and keywords from tweets related to IBD. The size of each keyword reflects the number of times it was mentioned by the patients—the larger the keyword, the more frequent it was.

The color of each keyword reflects the average sentiment it was given by the patients—green for positive sentiment, red for negative sentiment, and gray for neutral. Since an average sentiment of exact 0.0 is improbable, we considered all scores on the closed interval $[-0.05, 0.05]$ as neutral sentiments.

General words such as "people," "day," "time," and "thing" were frequent among all four groups and usually appeared in a negative context. However, the rest of the most frequent words were different between the three groups "All," "health and fitness," and "IBD-related," and the remaining group, "food and drink." The "food and drink" group captured nutrition-related words in positive sentiment.

Fig. 5 Word clouds of frequently mentioned keywords within patients' tweets by four groups: all tweets combined, tweets related to health and fitness, tweets related to food and drink, and tweets related to IBD. The keywords are colored based on their sentiments—green representing positive sentiment and red representing negative sentiment. The three clouds containing health-related keywords show a more negative sentiment than the one related to food and drink

In contrast, all the other three groups captured phrases related to IBD that mainly conveyed negative sentiment.

The independent samples t-tests between the average sentiments of the "food and drink" group and those of the other groups showed that the sentiment in the "food and drink" group was indeed more positive compared to the other groups (all three tests showed $p < 0.01$). No statistical differences were found between the average sentiments of the other three groups.

Looking at the clouds in Fig. 5, we can see that the "food and drink" cloud comprises relatively many green keywords, while the other three clouds are much redder. Words related to IBD are dominant in all three and appear in red. It means that patients tend to use a negative sentiment when they talk about their disease.

Emotion Classification

Patients and Control Group

The NLU module was used for emotion classification over tweets classified as "health and fitness" (53,598 patients' tweets and 3537 control group tweets) and tweets classified as "food and drink" (30,747 patients' tweets and 2,870 control group tweets). The module successfully detected emotions in 53,209 patients' tweets related to "health and fitness" (99.3%) and in 30,520 patients' tweets related to "food and drink" (99.3%). In the control group, it was able to detect emotions in 3525 tweets related to "health and fitness" (99.7%) and in 2859 tweets related to "food and drink" (99.6%). The remaining tweets were omitted from our dataset as they did not convey emotion according to the NLU module.

Each emotion's intensity of was averaged by a user and then again by category ("health and fitness" or "food and drink") within each group (patients vs. control). Significant differences were found between the groups in three emotions: sadness, joy, and fear, only in the "health and fitness" category. Patients' tweets related to health and fitness expressed more sadness and fear than those of the control group. The latter expressed more joy in their health-related tweets. Table 5 summarizes the results by showing the averaged intensity of each emotion and indicating the p value of the equivalent Wilcoxon test.

Figure 6 illustrates the differences in tail probabilities between the two groups within the "health and fitness"

Table 5 Means of averaged intensities by a user of the five emotions for each category and each group

Category	Group	Sadness	Joy	Fear	Disgust	Anger
Food and drink	Control	0.20511	0.44748	0.08030	0.09067	0.08856
	Patients	0.20949	0.46450	0.08072	0.07760	0.08653
Health and fitness	Control	0.28863	0.32710	0.13602	0.06818	0.09125
	Patients	0.30754**	0.29794*	0.14822**	0.05765	0.08156

Significant differences are marked with asterisks: * for p value smaller than 0.05 and ** for p value smaller than 0.01. The patients used more sadness and fear in their tweets, whereas the control group used more joy
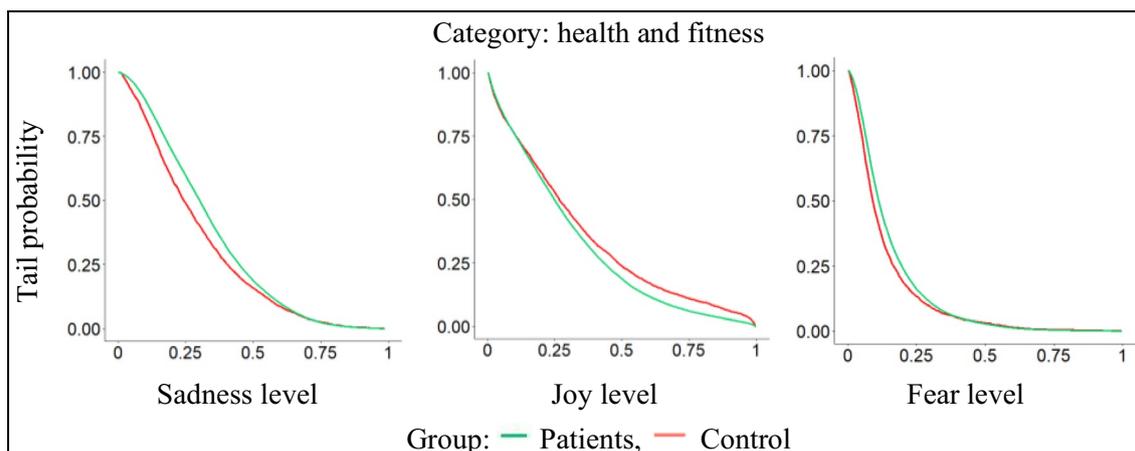

Category: health and fitness

Sadness level   Joy level   Fear level

Group: — Patients, — Control

Fig. 6 Comparison between the patients' tail probability (1 minus the cumulative probability) and the tail probability of the control group within the "health and fitness" category and the three emotions that showed significant differences: sadness, joy, and fear. For sadness and fear, the tail probabilities of the patients are higher than the ones of the control group, meaning that patients use more sadness and fear in their tweets. For joy, the situation is reversed

category and the three emotions that showed significant differences: sadness, joy, and fear. As seen in Fig. 6, the trends of the tail probabilities are in accordance with the results of the Wilcoxon tests. In both the sadness and fear graphs, the green lines representing patients' emotions are above the red lines representing the emotions of the control group. In the joy graph, however, the order is reversed, and the red line is always higher than the green line. It means that within the "health and fitness" category, the patients' tweets expressed more negative emotions (sadness and fear) and less positive ones, like joy, compared to the control group.

Discussion

Principal Results

In this research, we present a methodology for exploring patients' content on Twitter and comparing it with the content shared by other Twitter users. The study aimed to discover what patients with IBD are talking about on Twitter and to learn from the personal experience they share in their tweets.

In the first stage of the study, we constructed two datasets of tweets—the main dataset containing patients' tweets and a control dataset containing tweets from random Twitter users. Since demographic data are not available on Twitter, we suggested a method for sampling the Twitter population and forming the control group.

In the second stage of the study, the two collections of tweets were classified into thematic categories using the NLU module by IBM Cloud. It was found that the segmentation of topics discussed by IBD patients differs from the one discussed by the general population of Twitter users. While the random Twitter users frequently referred to politics and social affairs in their tweets, the IBD patients dedicated a considerable amount of their tweets (10%) to health and fitness. Only a small fraction (2.5%) of the patients' tweets related to religion or spirituality. This is a surprising finding since relaxing physical activities such as yoga, meditation, or mindfulness are considered helpful in maintaining the disease.

In the third stage of the study, keyword extraction with sentiment analysis was applied to all patients' tweets related to health, fitness, or nutrition to further investigate patients' discussions on Twitter. Using the NLU module, keywords from lifestyle-related tweets were identified and classified as positive, negative, or neutral to reflect on how patients deal with their disease. The analysis showed that the keywords were used with a more negative sentiment when the patients talked about their health. The sentiment of tweets related to nutrition was significantly higher. These findings

support a conclusion from a previous study determining that IBD patients usually express a negative sentiment when they talk about the disease and its symptoms but address more positively the diets and drugs that help manage them [40].

Unlike previous research related to patients' sentiment on Twitter [40, 51], we focused our research on entity sentiment rather than the sentiment of the entire tweet. The entity sentiment analysis can be used to characterize patients' sentiments toward the different treatments they try. Such an approach may uncover helpful lifestyles that help patients to cope with the disease.

In the fourth and final stage of the study, the NLU module was used for emotion classification over tweets classified as related to health or nutrition from both the patients' group and the control group. While no statistically significant differences were found between the groups in the "food and drink" category, interesting differences were found between the groups in the "health and fitness" category. Not only do patients tend to talk more about health than the general population, but they also tend to do so while expressing more sadness and fear and less joy. These findings also align with the keyword analysis, emphasizing how patients tend to use a more negative tone when discussing health. If random users talk about health in general while patients focus on their own disease, this might explain the different emotions each group expresses.

No statistically significant differences were found between the groups when expressing disgust or anger. However, these two emotions were less intense in general, in both groups and both categories. Hence, within the three dominant emotions (sadness, joy, and fear), we observed significant differences in how each group expresses themselves when addressing health-related issues.

This research suggests that there is room for collaboration between physicians and engineers regarding understanding chronic diseases. Due to the chronic nature of the disease and the fact that it involves bowel movements, IBD patients are compelled to follow special nutrition and maintain a calm routine. By collecting and analyzing patients' personal experiences on social media, we can monitor patients' lifestyles and support medical knowledge regarding IBD. We can identify and assess complementary treatments of diets and physical activity and maybe ease patients' processes of finding the right treatments. Although such analysis should not strive to replace doctors or draw conclusions of clinical nature, it may provide complementary recommendations for healthy lifestyles based on the wisdom of the crowd.

Limitations and Future Work

The NLU module by IBM Cloud was used in this research as a proof of concept, and we did not evaluate its results. Other NLP tools available in the market failed to return satisfactory

results for the category classification task, as they were oriented toward customer service reviews or could not cope with the short length of the tweets. Future research should consider evaluating the performance of the NLU module on Twitter data by manually classifying tweets into categories. Training designated algorithms on data from lifestyle-related tweets, like those used in this research, can benefit the presented analyses.

The control group in this research comprised only 57 random users, while the test group was larger and contained 377 IBD patients. Twitter Streaming API is free, and one can quickly collect many random users to increase the control group. However, the rest of the analysis—collecting all their tweets and using the NLU module to extract keywords from each one—is time-consuming and costly. In future work, we intend to continue the process and increase the control group to at least 100 users.

The timelines collection for the patients' group was done in November 2020, while the timelines collection for the control group was done in May 2021. The periodic gap between the collected data may affect the category segmentations. Another adjustment needed to balance the two groups is to collect the users' timelines over the same periods.

Our patients' dataset contained 337 patients that we unambiguously identified in our previous study. Applying the classifier of the prior study and identifying more patients with IBD can enrich the dataset and enable big data analysis. Future research should compare the results from such analysis with those presented in this study and evaluate the contribution of collecting more patients' data automatically. On the one hand, one can achieve a much larger dataset of patients. On the other hand, this dataset may contain erroneous identification of patients due to the imperfection of the classifier. It would be interesting to investigate how this trade-off affects the quality of the results.

We represented each user by their average values in the category and emotion analyses. We then used statistical tests to compare the patients' and the control groups' results. We considered all users equal once we obtained their averaged representations and did not consider the number of tweets they had in our dataset. Future research should try integrating weighted averages, for example, to take into consideration the frequency of tweets by a user.

Conclusions

This research presented a pipeline for analyzing patients' tweets and learning about their diseases from the content they share on Twitter. The study provided thematic analysis and keyword identification in tweets by patients with IBD.

The analysis showed that patients differ in the content they share on Twitter from other users and that the information they share can be used to derive insights regarding their disease. The methodology was detailed in four main stages to guide other researchers willing to reproduce the study or regenerate it for other medical conditions.

The research methods applied to IBD can also help explore other medical conditions. Research of other diseases involving strict dietary guidelines, like Celiac Disease or diabetes, can use the analysis of patients' tweets to better understand patients' difficulties with adhering to their new lifestyles. Research on diseases considered embarrassing, like HIV, can use such analysis to learn more about the constant struggle of patients living with the disease. The workflow presented in this study is feasibly extended to other chronic conditions.

Therefore, the contribution of this research is twofold. It provides an analytical contribution to text mining and social media and a practical contribution by better understanding chronic conditions and promoting a healthy lifestyle for chronic patients.

Data availability　The data that support the findings of this study are not openly available due to reasons of sensitivity and are available from the corresponding author upon reasonable request.

Declarations

Conflict of Interest　The authors declare that they have no conflict of interest.

Ethical Approval　The study was approved by the Human Subjects Research Committee of Ben-Gurion University of the Negev.

Informed Consent　Our study was based solely on publicly available Twitter data and did not perform any clinical intervention. The patients voluntarily provided all the information in their tweets or profile description. To adhere to ethical norms and maintain user privacy, we only publish aggregated results that do not reveal the specific users. The examples containing direct quotes from tweets are presented in this study after obtaining informed consent from their authors.

References

1. Kaplan GG. The global burden of IBD: from 2015 to 2025. Nat Rev Gastroenterol Hepatol. 2015;12(12):720–7. https://doi.org/10.1038/nrgastro.2015.150.
2. Loftus EV Jr. Clinical epidemiology of inflammatory bowel disease: incidence, prevalence, and environmental influences. Gastroenterology. 2004;126(6):1504–17. https://doi.org/10.1053/j.gastro.2004.01.063.

3. Roccetti M, Marfia G, Salomoni P, Prandi C, Zagari RM, Gningaye Kengni FL, et al. Attitudes of Crohn's disease patients: infodemiology case study and sentiment analysis of Facebook and Twitter posts. JMIR Public Health Surveill. 2017;3(3): e51. https://doi.org/10.2196/publichealth.7004.

4. Trivedi I, Keefer L. The emerging adult with inflammatory bowel disease: challenges and recommendations for the adult gastroenterologist. Gastroenterol Res Pract. 2015;2015:260807. https://doi.org/10.1155/2015/260807.

5. Norton BA, Thomas T, Lomax KG, Dudley-Brown S. Patient perspectives on the impact of Crohn's disease: results from group interviews. Patient Prefer Adherence. 2012;6:509–20. https://doi.org/10.2147/PPA.S32690.

6. Rubin DT, Dubinsky MC, Panaccione R, Siegel CA, Binion DG, Kane SV, et al. The impact of ulcerative colitis on patients' lives compared to other chronic diseases: a patient survey. Dig Dis Sci. 2010;55(4):1044–52. https://doi.org/10.1007/s10620-009-0953-7.

7. Devlen J, Beusterien K, Yen L, Ahmed A, Cheifetz AS, Moss AC. The burden of inflammatory bowel disease: a patient-reported qualitative analysis and development of a conceptual model. Inflamm Bowel Dis. 2014;20(3):545–52. https://doi.org/10.1097/01.mib.0000440983.86659.81.

8. Brydolf M, Segesten K. Living with ulcerative colitis: experiences of adolescents and young adults. J Adv Nurs. 1996;23(1):39–47. https://doi.org/10.1111/j.1365-2648.1996.tb03133.x.

9. Hall NJ, Rubin GP, Dougall A, Hungin AP, Neely J. The fight for 'health-related normality': a qualitative study of the experiences of individuals living with established inflammatory bowel disease (ibd). J Health Psychol. 2005;10(3):443–55. https://doi.org/10.1177/1359105305051433.

10. De Choudhury M, Morris MR, White RW. Seeking and sharing health information online: comparing search engines and social media. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2014 Presented at: CHI '14;, 2014; Toronto, Canada, p. 1365–1376. https://doi.org/10.1145/2556288.2557214.

11. Becker KL. Cyberhugs: creating a voice for chronic pain sufferers through technology. Cyberpsychol Behav Soc Netw. 2013;16(2):123–6. https://doi.org/10.1089/cyber.2012.0361.

12. Wiese J, Kelley PG, Cranor LF, Dabbish L, Hong JI, Zimmerman J. Are you close with me? Are you nearby?: investigating social groups, closeness, and willingness to share. In: Proceedings of the 13th International Conference on Ubiquitous Computing. 2011 Presented at: UbiComp '11; September 17–21, 2011; Beijing, China, p. 197–206. https://doi.org/10.1145/2030112.2030140.

13. Paek HJ, Hove T, Ju Jeong H, Kim M. Peer or expert? The persuasive impact of YouTube public service announcement producers. Int J Advert. 2015;30(1):161–88. https://doi.org/10.2501/IJA-30-1-161-188.

14. Lin WY, Zhang X, Song H, Omori K. Health information seeking in the Web 2.0 age: trust in social media, uncertainty reduction, and self-disclosure. Comput Human Behav. 2016;56:289–94. https://doi.org/10.1016/j.chb.2015.11.055.

15. Frohlich DO. The social construction of inflammatory bowel disease using social media technologies. Health Commun. 2016;31(11):1412–20. https://doi.org/10.1080/10410236.2015.1077690.

16. Macdonald GG, Koehn C, Attara G, Stordy A, Allerdings M, Leese J, et al. Patient perspectives on the challenges and responsibilities of living with chronic inflammatory diseases: qualitative study. J Participatory Med. 2018;10(4): e10815. https://doi.org/10.2196/10815.

17. Khan A, Silverman A, Rowe A, Rowe S, Tick M, Testa S, et al. Who is saying what about Inflammatory Bowel Disease on Twitter? In: GW Annual Research Days 2018. 2018 Presented at: GW Research '18; April 10–11, 2018; Virtual p. 176 URL: https://hsrc.himmelfarb.gwu.edu/gw_research_days/2018/SMHS/62/.

18. Rowe A, Rowe S, Silverman A, Borum ML. P024 Crohn's disease messaging on Twitter: who's talking? Gastroenterology. 2018;154(1):S13–4. https://doi.org/10.1053/j.gastro.2017.11.059.

19. O'Neill P, Shandro B, Poullis A. Patient perspectives on social-media-delivered telemedicine for inflammatory bowel disease. Future Healthc J. 2020;7(3):241–4. https://doi.org/10.7861/fhj.2020-0094.

20. Frohlich DO, Zmyslinski-Seelig AN. How uncover Ostomy challenges ostomy stigma, and encourages others to do the same. New Media Soc. 2014;18(2):220–38. https://doi.org/10.1177/2F1461444814541943.

21. Stemmer M, Parmet Y, Ravid G. Identifying Patients With Inflammatory Bowel Disease on Twitter and Learning From Their Personal Experience: Retrospective Cohort Study. J Med Internet Res. 2022;24(8): e29186. https://doi.org/10.2196/29186.

22. Twitter API Documentation. Twitter I. 2022. URL: https://developer.twitter.com/en/docs/twitter-api [Accessed 1 Mar 2022].

23. Natural Language Understanding. IBM Cloud API Docs. 2020. URL: https://cloud.ibm.com/apidocs/natural-language-understanding [Accessed 15 July 2021].

24. Pang B, Lee L. Opinion mining and sentiment analysis. Found Trends® Inf Retr. 2008;2(1–2):1–135. https://doi.org/10.1561/1500000011.

25. Karami A, Dahl AA, Turner-McGrievy G, Kharrazi H, Shaw G Jr. Characterizing diabetes, diet, exercise, and obesity comments on Twitter. Int J Inf Manage. 2018;38(1):1–6. https://doi.org/10.1016/j.ijinfomgt.2017.08.002.

26. Kausar MA, Soosaimanickam A, Nasar M. Public sentiment analysis on Twitter data during COVID-19 outbreak. Int J Adv Comput Sci Appl. 2021. https://doi.org/10.2196/21978.

27. Thelwall M, Thelwall S. A thematic analysis of highly retweeted early COVID-19 tweets: consensus, information, dissent and lockdown life. Aslib J Inf Manag. 2020. https://doi.org/10.1108/AJIM-05-2020-0134.

28. Grewal US, Gupta A, Doggett J, Lou E, Gusani NJ, Maitra A, et al. Twitter conversations about pancreatic cancer by health care providers and the general public: thematic analysis. JMIR cancer. 2022;8(1): e31388. https://doi.org/10.2196/31388.

29. Rohde JA, Sibley AL, Noar SM. Topics analysis of reddit and twitter posts discussing inflammatory bowel disease and distress from 2017 to 2019. Crohns Colitis 360. 2021;3(3):otab044. https://doi.org/10.1093/crocol/otab044.

30. Facanali MR, Bortolozzo Graciolli Facanali C, Queiroz NSF, Sobrado CW, Nahas SC, Safatle-Ribeiro AV. # Crohn's: historical cohort of twitter activity. Crohn's Colitis 360. 2021;3(1):otaa075. https://doi.org/10.1093/crocol/otaa075.

31. Kamath P, Kursewicz C, Ingrasci G, Jacobs R, Agarwal N, Nouri K. Analysis of patient perceptions of Mohs surgery on social media platforms. Arch Dermatol Res. 2019;311(9):731–4. https://doi.org/10.1007/s00403-019-01944-7.

32. Gabarron E, Larbi D, Dorronzoro E, Hasvold PE, Wynn R, Årsand E. Factors engaging users of diabetes social media channels on facebook, twitter, and instagram: observational study. J Med Internet Res. 2020;22(9): e21204. https://doi.org/10.2196/21204.

33. Thomas J, Prabhu AV, Heron DE, Beriwal S. Twitter and brachytherapy: An analysis of "tweets" over six years by patients and health care professionals. Brachytherapy. 2018;17(6):1004–10. https://doi.org/10.1016/j.brachy.2018.07.015.

34. Cheng TY, Liu L, Woo BK. Analyzing Twitter as a platform for Alzheimer-related dementia awareness: thematic analyses of tweets. JMIR aging. 2018;1(2): e11542. https://doi.org/10.2196/11542.

35. Mullins CF, Ffrench-O'Carroll R, Lane J, O'Connor T. Sharing the pain: an observational analysis of Twitter and pain in Ireland. Reg Anesth Pain Med. 2020;45(8):597–602. https://doi.org/10.1136/rapm-2020-101547.

36. Zhang Y, Lyu H, Liu Y, Zhang X, Wang Y, Luo J. Monitoring depression trends on twitter during the COVID-19 pandemic: observational study. JMIR Infodemiol. 2021;1(1): e26769. https://doi.org/10.2196/26769.

37. Chen X, Sykora MD, Jackson TW, Elayan S. What about mood swings: Identifying depression on twitter with temporal measures of emotions. Companion Proceedings of the The Web Conference 2018; 2018. https://doi.org/10.1145/3184558.3191624.

38. Suhasini M, Badugu S. Two step approach for emotion detection on twitter data. Int J Comput Appl. 2018;179(53):12–9. https://doi.org/10.5120/ijca2018917350.

39. Suhasini M, Srinivasu B. Emotion detection framework for twitter data using supervised classifiers. In: Data engineering and communication technology. Springer; 2020. p. 565–76. https://doi.org/10.1007/978-981-15-1097-7_47.

40. Pérez-Pérez M, Pérez-Rodríguez G, Fdez-Riverola F, Lourenço A. Using twitter to understand the human bowel disease community: exploratory analysis of key topics. J Med Internet Res. 2019;21(8): e12610. https://doi.org/10.2196/12610.

41. Stemmer M, Ravid G, Parmet Y. A Framework for Identifying Patients on Twitter and Learning from Their Personal Experience. 2020. In Proceedings of the 28th European Conference on Information Systems (ECIS), An Online AIS Conference, June 15–17, 2020. https://aisel.aisnet.org/ecis2020_rp/194.

42. Gabarron E, Dorronzoro E, Rivera-Romero O, Wynn R. Diabetes on Twitter: a sentiment analysis. J Diabetes Sci Technol. 2019;13(3):439–44. https://doi.org/10.1177/1932296818811679.

43. Sampled stream. Twitter I. 2021. URL: https://developer.twitter.com/en/docs/twitter-api/tweets/sampled-stream/api-reference/get-tweets-sample-stream [Accessed 1 Mar 2022].

44. Maier M. Dirichlet Regression in R. Version 0.4–0. R Foundation for Statistical Computing, Vienna, Austria 2012.

45. Maier MJ. DirichletReg: Dirichlet regression for compositional data in R. 2014.

46. Maier MJ, Maier MMJ. Package 'DirichletReg'. 2015.

47. Lewis F, Butler A, Gilbert L. A unified approach to model selection using the likelihood ratio test. Methods Ecol Evol. 2011;2(2):155–62. https://doi.org/10.1111/j.2041-210X.2010.00063.x.

48. Kuha J. AIC and BIC: comparisons of assumptions and performance. Sociol Methods Res. 2004;33(2):188–229. https://doi.org/10.1177/2F0049124103262065.

49. Burnham KP, Anderson DR. Multimodel inference: understanding AIC and BIC in model selection. Sociol Methods Res. 2004;33(2):261–304. https://doi.org/10.1177/2F0049124104268644.

50. MacFarland TW, Yates JM. Mann-Whitney *U* test. In: Introduction to nonparametric statistics for the biological sciences using R. Springer; 2016. p. 103–32. https://doi.org/10.1007/978-3-319-30634-6_4.

51. Sailunaz K, Alhajj R. Emotion and sentiment analysis from Twitter text. J Comput Sci. 2019;36:101003. https://doi.org/10.1016/j.jocs.2019.05.009.