Natural language processing of online support group postings reveals patients' perspectives on strategies for managing psoriasis

Karan Rajalingam,[1] Phillip M. Johansen,[1] Nicole J. Levin,[1] Jerry Qi,[1] Oge Marques[2]

[1]Charles E. Schmidt College of Medicine, Florida Atlantic University, Boca Raton, FL; [2]Department of Computer and Electrical Engineering and Computer Science, Florida Atlantic University, Boca Raton, FL, United States

Correspondence: Karan Rajalingam, Florida Atlantic University, Charles E. Schmidt College of Medicine, 980 N University Dr, Box 599, Boca Raton, FL 33431, United States.
Tel.: +1.3103835879.
E-mail: krajalingam2020@health.fau.edu

Abstract

Psoriasis is a chronic skin disorder, and patients encounter high physical and psychosocial burdens. Social media forums feature extensive patient-generated comments. We hypothesized that analyzing patient-posted comments using natural language processing would provide insights into patient engagements, sentiments, concerns, and support, which are vital for the holistic management of psoriasis. We collected 32,000 active user comments posted on Reddit. We applied Latent Dirichlet Allocation to categorize posts into popular topics and employed spectral clustering to establish cohesive themes and word representation frequency within these topics. We sorted posts into 29 significant topics of discussion and categorized them into four categories: management (37.48%), emotion (21.57%), presentation (19.79%), and others (3.57%). The frequent posts on management were diet (7.23%), biologics (6.95%), and adverse effects (3.88%). The emotion category comprised negative sentiments (11.02%), encouragement (5.49%), and gratitude (5.06%). The presentation topic included a discussion of scalp (5.69%), flare-timing (3.63%), and arthritis (2.64%). Others comprised differential diagnosis (5.01%), leaky gut (4.12%), and referrals (3.70%). This study identified patients' experiences and perspectives associated with psoriasis, which should be considered to tailor support systems to improve their quality of life.

Introduction

Psoriasis is a chronic, immune-mediated, inflammatory skin disease characterized by red scaly plaques on any skin surface. It is commonly observed on the elbows, knees, scalp, and lower back.[1] According to the World Psoriasis Day Consortium, over 125 million people worldwide have psoriasis. Patients with psoriasis may have significant comorbidities, including cardiovascular disease and psoriatic arthritis.[2,3] Additionally, they may feel tremendous stigmatization, leading to anxiety and depression.[4,5] Despite advances in psoriasis treatment, a support system for patient education to improve quality of life remains unmet.[6,7]

Given easy access and anonymity, patients commonly research social media websites for medical conditions before seeking professional expertise.[8] With the immense negative physical and social impacts of psoriasis, patients turn to social media to acquire health information and improve their quality of life.[9] In a cross-sectional questionnaire-based study, Swiss researchers observed that 82.4% of patients with dermatological diseases use the Internet to access medical information.[10] Additionally, they have employed social media support groups to exchange information about disease management and provide one another with emotional support.[11] Compared with face-to-face conversations with physicians, social media provides an anonymous environ-

ment that may facilitate a more open disclosure of symptoms experienced, remedies attempted, personal concerns, or controversial ideas. However, there is a growing concern that these outlets are susceptible to misinformation that may be transmitted inadvertently, and influence negative attitudes toward evidence-based treatment.[12] The utility of social media for patients with psoriasis and its positive effects on them are not entirely understood.

Reddit is the world's 7th most popular social media site in the world with 430 million active monthly users. In recent years, psoriasis-related discussions have grown considerably, with memberships exceeding 32,000 people. Reddit is a unique community platform that allows users to preserve their anonymity; therefore, users may feel more comfortable sharing their personal experiences and issues. Furthermore, users can pose open-ended questions related to psoriasis and receive responses from other members of the forum. Natural language processing (NLP) is a branch of artificial intelligence that enables computers to interpret narrative medical records and massive text messages to extract insights.[13] We hypothesized that the NLP of user-generated social media posts would provide actionable insights into the dermatologic conditions and therapies that patients with psoriasis undergo.

Materials and Methods

Reddit forums are composed of chronologically sorted initial posts and response posts which are referred to as submissions and comments respectively. We scraped all posts and comments over the five years from January 1st, 2017, to January 1st, 2022, from the psoriasis-specific Reddit forum accessed through the link https://www.reddit.com/r/psoriasis. All subsequent analyses were performed using Python libraries. As the analysis was performed on data extracted from a publicly accessible forum, this study was exempt from obtaining informed consent and institutional review board approval.[14]

A total of 15,741 submissions and 153,990 comments from 169,731 posts were included for analysis (Figure 1). Web scraping and post aggregation were performed using the Python libraries json and pandas. We excluded 6,769 duplicate posts, defined as those with the same content associated with the same username, and filtered out posts automatically generated by Internet bots. Subsequently, we searched the content of the remaining 162,962 posts to extract posting dates and hyperlinks.

Afterward, the posts were preprocessed using NLP, a field of artificial intelligence. Posts less than 20 characters in length (n=10,327) were discarded to exclude terse posts from the analysis. The remaining 152,635 posts were stripped of punctuation and the individual words were lemmatized into their root form. This preprocessing step was executed using the Python ntlk library.[15]

Finally, Latent Dirichlet Allocation (LDA), an unsupervised machine learning algorithm, was applied to classify the documents. LDA uses the frequency of word occurrences within posts to associate them with similar posts to constitute a group. The number of groups required for the LDA to split the posts was empirically determined. We adjusted the number of groups stepwise, starting with five groups and adding five in each step until we reached an interpretable consensus that did not incur a significant overlap. This analysis was performed using the Python Natural Language Toolkit and scikit-learn libraries.[15]

We extracted two lists of representative word sets from each LDA post grouping to assign labels. The first set consisted of the most frequently occurring words in each group. The second set contained the most relevant words for each group. The term frequency-inverse document frequency (TF-IDF) algorithm generates the most relevant words. In TF-IDF, the relevance of words is weighted directly by their frequency and inversely by their coexistence with other groups. Each group was subsequently assigned a topic label based on the words within each set and topics with similar labels were combined. Further, these topics were organized into more significant categories, and the frequencies of posts within each topic and category were tabulated.

Results

There was consistent growth in forum activity over the study period, starting with 6,846 posts in the first year and increasing to 136,716 posts in the last year of the study period. There were 1,666 specific hyperlinks that were shared 11,152 times among users over the study period (Figure 2). The most frequently shared websites were reddit.com (n=3426), nih.gov (n=1004), amazon.com (n=754), imgur.com (n=622), YouTube (n=396), Wikipedia (n=247), and psoriaisis.org (n=214). Reddit.com links directed queries to other forum posts, including related personal anecdotes and central hub pages for frequently asked questions. The nih.gov link was routed to citations of evidence-based
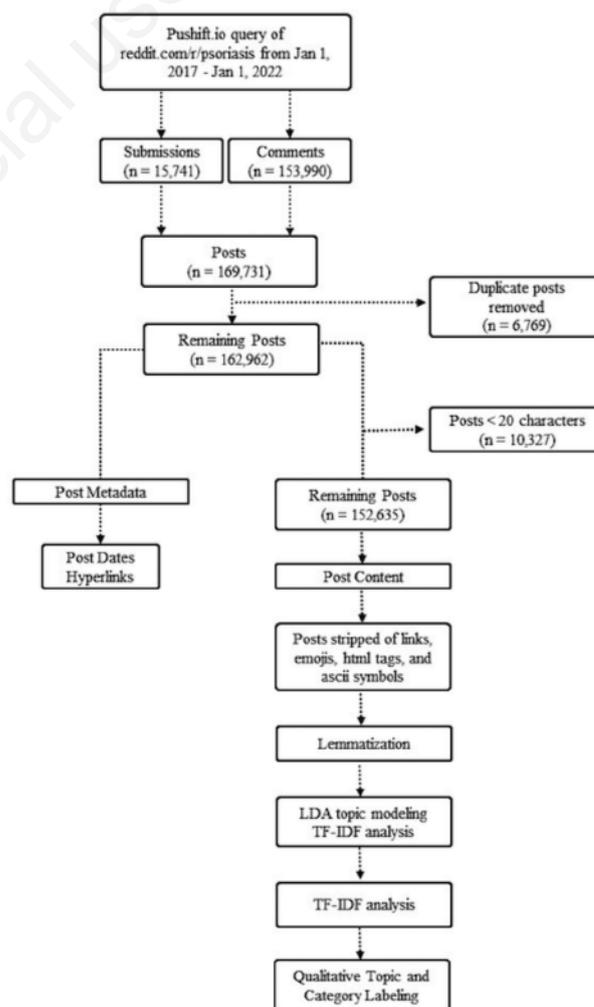

Figure 1. Forum analysis pipeline.

research. Amazon.com links were employed for product recommendations, such as moisturizers and cotton garments. Imgur.com is an image-sharing platform that is utilized for uploading pictures of psoriatic skin lesions. The remaining hyperlinks were directed towards various destinations, including personal blogs, dermatology practice webpages, international websites, and specialty product websites.

LDA was applied to divide the posts into topics. A stepwise trial-and-error methodology determined that executing the algorithm with 40 topics yielded the most diverse and interpretable consensus. Word sets were sampled corresponding to the most frequently used and relevant words for each topic. Further, they were used to assign labels to topics manually. Nine topics were labeled and combined, however, two topics comprising 6,926 posts, were omitted because they could not be assigned labels. We identified 29 unique topics that were labeled into broader categories. Table 1 lists these topics and their categories. An abridged word set consists of the highest frequency words occurring in each topic or the most relevant words determined by the TF-IDF analysis.

These 145,708 posts were divided into 29 topics and grouped into four categories. The categories were management (37.48%), emotion (21.57%), presentation (19.79%), and others (3.57%). The most frequently discussed topics in management were diet (7.23%), biology (6.95%), and adverse effects (3.88%). The posts in this category comprised user discussions on anecdotal tips and personal experiences regarding psoriasis management. The emotion category included negativity (11.02%), encouragement (5.49%), and gratitude (5.06%). Posts grouped into this category involved emotional venting about the psychosocial impact of the

disease on personal lives, positive comments on encouragement, and gratitude for well-received management advice. The presentation category comprised discussions of the scalp (5.69%), flare timing (3.63%), and arthritis (2.64%). Descriptive posts on current or prior flares were sorted into this category. The other category comprised a collection of unrelated topics including differential diagnoses (5.01%), leaky gut (4.12%), and referrals (3.70%). Examples of posts on topics in each category are listed in Table 2.
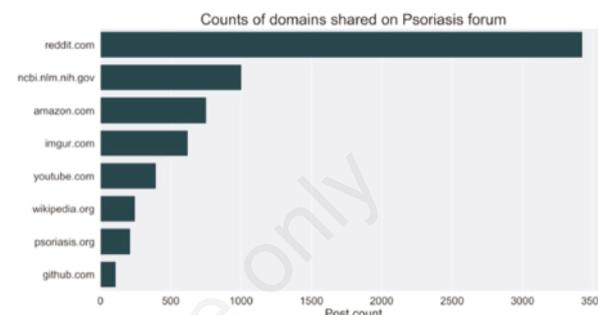

Figure 2. Tally of website domains shared on psoriasis reddit forum.

Table 1. Summary of categories, topics, and their relative frequencies.

Category	Topic	Posts	%	Word set
Management 37.48%	Diet	10,538	7.23	Gut, carb, fruit, elimination, keto, nightshade, vegetable, egg, aip, fiber, bread
	Biologic	10,121	6.95	Humira, insurance, cosentyx, taltz, biologic, stelara, pay, work, switch, skyrizi
	Adverse effects	5,652	3.88	Loading, nausea, headache, fatigue, starter, gi, excited, diarrhea, maintenance
	Steroid	5,221	3.58	Cream, use, steroid, buy, help, try, just, psoriasis, work, lotion, really, area
	Moisturizer	4,411	3.03	Cerave, rinse, aveeno, gently, eucerin, descale, moisturizing, gentle, fragrance
	Vitamin D	3,707	2.54	Calcipotriene, atrophy, potent, calcitriol, keratolytic, thinning, potency, taclonex
	Haircut	3,101	2.13	Bald, dye, hairline, buzz, curly, greasy, comb, barber, stylist, wig, haircut, hat
	Methotrexate	2,862	1.96	Nausea, fatigue, folic, mood, diarrhea, injector, mtx, starter, acitretin, flu, adverse
	Makeup	2,325	1.60	Makeup, cheek, bright, foundation, concealer, peel, corner, sunscreen, snail, gentle
	Sun exposure	2,362	1.62	Sunscreen, deficiency, stats, calcium, deficient, beach, bulb, intake, multivitamin
	Salt bath	2,258	1.55	Salt, bath, answer, wear, easy, time, just, make, oatmeal, mineral
	Light therapy	2,060	1.41	Uva, bulb, narrowband, nm, wand, device, wavelength, handheld, machine, emit
Emotion 21.57%	Negativity	16,050	11.02	Conscious, esteem, sex, mentally, gross, embarrassed, vent, cope, social, ashamed
	Encouragement	8,004	5.49	Good, luck, wish, result, glad, hear, hope, work, thing, great, pretty, idea
	Gratitude	7,368	5.06	Sharing, insight, ah, informative, message, encouraging, detailed, sweet, reassuring
Presentation 19.79%	Scalp	8,287	5.69	Conditioner, comb, rinse, greasy, shea, castor, vinegar, jojoba, tgel, essential
	Flare timing	5,291	3.63	Year, away, time, month, start, stop, come, week, flare, derm, itch
	Arthritis	3,844	2.64	Psa, hip, ra, mri, ache, wrist, tendon, rapharub, stiffness, swollen, swelling, fibro
	Guttate	3,560	2.44	Tonsil, tonsillectomy, stressful, guttate, pustular, college, gut, ent, pyogenes
	Plaque	2,788	1.91	Amp, leg, area, arm, elbow, patch, scalp, small, psoriasis, spot, body, knee
	Ear	2,587	1.78	Canal, wax, hearing, ent, flu, plug, drum, headphone, earwax, gross, earring
	Nail	1,469	1.01	Pitting, pit, polish, cuticle, toenail, trim, fingernail, file, cross, paint, ridge, acrylic
	Palmoplantar	1,009	0.69	Hand, foot, product, heal, psoriasis, use, mention, just, experience, flare, try, skin
Other 21.16%	Differential Dx	7,307	5.01	Pustular, reliably, misdiagnose, description, seborrheic, ringworm, bite, pityriasis
	Leaky gut	6,009	4.12	Gut, leaky, microbiome, genetic, scientific, permeability, intestinal, syndrome, gene
	Referral	5,397	3.70	Doctor, say, tell, dermatologist, ask, prescribe, just, appointment, time, talk
	Insurance	4,725	3.24	Copay, assistance, card, healthcare, deny, deductible, appeal, private, referral
	Diagnostics	3,964	2.72	Symptoms, ana, crp, nbsp, antibody, lupus, tb, gut, report, marker, donate
	Pathophysiology	3,431	2.35	Psoriasis, infection, risk, immune, gt, study, drug, patient, increase, cancer

Mtx, methotrexate; gi, gastrointestinal; nm, nanometer; psa, antinuclear antibody; ra, rheumatoid arthritis; mri, magnetic resonance imaging; ent; ear, nose, and throat; ana, antinuclear antibody; crp, c-reactive protein; tb, tuberculosis; nbsp, non-breaking space; gt, geographic tongue.

Discussion

The results indicate a broad range of topics discussed by patients with psoriasis in online forums. The year-on-year increase in post-activity signifies the importance of this forum as a valuable resource for patients. The large number of posts made in the last five years has allowed us to quantify topics of discussion that yield insights into the top concerns of patients with psoriasis. There is a high degree of interest in non-pharmacological management. Among the medical management options, biologics comprise the bulk of the discussion. This forum is effectively used as an emotional venting platform, and positive comments are made to support other users. Diet is the most widely discussed non-pharmacological management strategy. A search for dietary suggestions within the forum yields a discussion of popular dietary regimens that users anecdotally associate with either remission or symptom worsening. The example post for diet in Table 2 mentions the word "Pagano," a commonly recommended diet that purports to be a "drug-free program" to manage psoriasis. Additionally, interest is observed in the influence of the microbiome on psoriasis, as demonstrated by Leaky Gut. This highlights the need for randomized controlled studies on the influence of diet and the microbiome on psoriasis that may provide evidence-based guidelines regarding the efficacy of alternative management strategies.

The second most frequently discussed management strategy is the use of biologic therapies. A previous study based on 5,524 Twitter posts on psoriasis rendered over two weeks in December 2016 revealed that 99 (1.79%) posts discussed biological medications.[16] Our study revealed a marked increase (6.95%) in the discussion of these agents, reflecting a growing consideration of these treatment options. Additionally, there were sizable discussions regarding adverse effects (3.88%) and insurance (3.24%). A recent report on 1141 posts discussing biologics extracted from Reddit revealed that 61.3% had positive sentiments about biologics and their efficacy, 24.0% were neutral, and 14.7% were negative.[17] Our study further disclosed that newer agents discussed in these forums include risankizumab (Skyrizi), tildrakizumab (Ilumya), and guselkumab (Tremfya). Previous studies on other biologics have suggested that forums can be used to sense the adverse events of newly released biologics well before official Food and Drug Administration (FDA) reports and that the same strategy may be applied to this forum.[18-20]

Additionally, our results revealed a significant amount of emotional discussions. This included narratives of personal hardships and social limitations that psoriasis imposes on occupational and interpersonal relationships. These reports reflect the high percentage of psychiatric comorbidities in patients with psoriasis.[3] A global survey of 1678 patients with psoriasis from 11 countries unveiled that 84% of patients experienced stigma and discrimination and reported a negative impact of their disease on work, emotions, and relationships. While our analysis disclosed a preponderance of negative sentiment posts (11.02%), it also showed a comparable number of posts expressing encouragement and gratitude (10.55%). This presents evidence that patients rely on emotional support in this forum. Many emotional posts reiterated the responsibility of the provider to assess psychological needs and refer patients for appropriate help. Consistent with our findings, an early exploratory analysis of 250 patients recruited from 5 online support groups demonstrated that virtual communities offered users psychological and social support.[21] Previous studies examining YouTube as a source of information on psoriasis treatment have raised concerns about the spreading of misinformation.[22-24] Patient testimonials endorsing alternative medicine and discouraging evidence-based medicine received disproportionately high levels of engagement. Our analysis uncovered topics and keywords that did not align with evidence-based management strategies. For example, a few users created posts claiming that tonsillectomy could cure specific subtypes of psoriasis; however, no studies supported their theories.

Evidence suggests that this forum engages in practices against the spread of misinformation. The results of our study showed strong patient engagement with a varied spectrum of treatment modalities, indicating that patients in this forum are invested in experimenting with all modalities to improve their disease outcomes. Forum moderators also surveyed posts to monitor submissions and ensured that there were no posts promoting questionable, non-FDA-approved miracle cures that may be unsafe. Additionally, a sizable number of link-sharing studies are directed to evidence-based medicine articles. Finally, the referral (3.70%) topic consisted of posts encouraging users to seek professional medical care from dermatologists.

The study has a few limitations. LDA is an unsupervised machine-learning model that classifies posts into topics based on the frequency of words within the posts. Thus, the classification serves only as an estimate of the extent of topics under discussion. However, as the Reddit forum is anonymous, patient data are viable and useful for dermatological research and public engagement, especially for common dermatology topics such as tanning, acne, and psoriasis. Patients with psoriasis have various options for interacting with social media forums. We were limited to Reddit.com users, who tend to be young men. By contrast, psoriasis typically affects men and women of all ages. Therefore, the data may not be generalizable to all patients with psoriasis.

Table 2. Example posts from 4 topics.

Diet	Biologics
"My scalp has been the most stubborn of it all. I have done experiments in the past by reducing alcohol or following diets like Dr. Pagano. The Pagano diet cleared up things fairly quickly all over for me but it was so hard to sustain."	"After nearly 10 years having 90% of my body covered with psoriasis, I had my first dose of Skyrizi 2 days ago. Seeing all the positive comments on how good it is, I have great hope that this will be life-changing. Took some before pics so will put them up here once I see some good change"
Negativity	**Encouragement**
"I get stressed about my skin and it seems to get worse even when I think I'm doing the right thing. I know some of it is in my head but it is so demotivational… The embarrassment and awkwardness to hide my skin eats me up daily, I'm not who I was socially 10 years ago and I'm afraid I'm going down a path of seclusion sometimes. I do have a few people to confide with but they don't really know the struggle."	"Hey mate, everything you're feeling is perfectly valid and I think we all agree these feelings occur more often than any of us would like. Just know you're more than your skin, what matters is who you are as a person. Send me a DM if you want someone to vent or talk to. I'm always open to discussion and helping out. Hang in there friend."

Another limitation is that not all posting users have psoriasis, whereby a portion of users post asking whether they have psoriasis based on symptomatology, which may skew the distribution of the results. Furthermore, the LDA methodology divides posts into topics based on the frequencies of words within posts but without an inherent understanding of the context or the English language. In addition, this machine learning model is unsupervised, which indicates that the model outputs have no standard of comparison. Therefore, our findings warrant further investigation in prospective trials.

Conclusions

Machine learning analysis of patient posts on the Reddit forum provided insights into contemporary trends in psoriasis management and the unmet needs of the psoriasis patient population. Further, it reveals the potential benefits and drawbacks of social media use among patients. This analysis highlights the need for the inclusion of dialogue and the perspective of patients. It drives shared decision-making in optimal disease management, including enhanced access to psychological support. Patients may omit information about their routine care or psychological burden during visits but discuss these aspects of psoriasis in online forums. Clinicians must remain cognizant of this dilemma and keep pace with evolving management options so that patients can receive the best care for their physical and emotional well-being.

Despite being an anonymous forum, users support each other through positive statements and explore alternative therapies with measured scrutiny. However, information seekers may be misled by unsound medical advice. With the rapid growth of the forum in recent years and an estimated 16 million people affected by psoriasis in the U.S., it has become critical for medical institutions to relay a presence in this forum and other online spaces to ensure that patients' perspectives are perceived. These findings underscore the need for patient education programs, understanding patients' quality of life in physical and mental health spaces, productive conversations, and shared decision-making to improve strategies for enhancing the overall quality of life in patients.

References

1. Griffiths CE, Barker JN. Pathogenesis and clinical features of psoriasis. Lancet 2007;370:263-71.
2. Chandran V, Raychaudhuri SP. Geoepidemiology and environmental factors of psoriasis and psoriatic arthritis. J Autoimmun 2010;34:J314-21.
3. Yeung H, Takeshita J, Mehta NN, et al. Psoriasis severity and the prevalence of major medical comorbidity: a population-based study. JAMA Dermatol 2013;149:1173-9.
4. Rahman SM, Abduelmula A, Jafferany M. Psychopathological symptoms in dermatology: A basic approach toward psychocutaneous disorders. Int J Dermatol 2023;62:346-56.
5. Patel PM, Sanchez-Melendez SN, Nambudiri VE. A narrative review of studies assessing the quality of life in patients with generalized pustular psoriasis. Exp Dermatol 2023;32:1227-34.
6. Lebwohl MG, Kavanaugh A, Armstrong AW, Van Voorhees AS. US Perspectives in the Management of Psoriasis and Psoriatic Arthritis: Patient and Physician Results from the Population-Based Multinational Assessment of Psoriasis and Psoriatic Arthritis (MAPP) Survey. Am J Clin Dermatol 2016;17:87-97.
7. Lebwohl MG, Bachelez H, Barker J, et al. Patient perspectives in the management of psoriasis: results from the population-based Multinational Assessment of Psoriasis and Psoriatic Arthritis Survey. J Am Acad Dermatol 2014;70:871-81.e1-30.
8. Savas JA, Huang KE, Tuchayi SM, Feldman SR. Understanding the influence of social media in medicine: lesson learned from Facebook. Dermatol Online J 2014;20.
9. DeBord LC, Patel V, Braun TL, Dao HJr. Social media in dermatology: clinical relevance, academic value, and trends across platforms. J Dermatolog Treat 2019;30:511-8.
10. Gantenbein L, Navarini AA, Maul LV, et al. Internet and social media use in dermatology patients: Search behavior and impact on patient-physician relationship. Dermatol Ther 2020;33:e14098.
11. Amir M, Sampson BP, Endly D, et al. Social networking sites: emerging and essential tools for communication in dermatology. JAMA Dermatol 2014;150:56-60.
12. Qi J, Trang T, Doong J, et al. Misinformation is prevalent in psoriasis-related YouTube videos. Dermatol Online J 2016;22.
13. Harrison CJ, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction to natural language processing. BMC Med Res Methodol 2021;21:158.
14. Moreno MA, Goniu N, Moreno PS, Diekema D. Ethics of social media research: common concerns and practical considerations. Cyberpsychol Behav Soc Netw 2013;16:708-13.
15. Abraham A, Pedregosa F, Eickenberg M, et al. Machine learning for neuroimaging with scikit-learn. Front Neuroinform 2014;8:14.
16. Menzies S, Daly S, McKenna DB. Social media and psoriasis treatment: what are people saying on Twitter? Br J Dermatol 2019;180:1527-8.
17. Maynard N, Yee D, Carranza C, et al. Evaluating Sentiment, Engagement, and Perception of Biologics Among Patients with Psoriasis and Psoriatic Arthritis on Reddit. J Drugs Dermatol 2023;22:306-9.
18. Zhang W, Mukerjee S, Qin H. Topics and Sentiments Influence Likes: A Study of Facebook Public Pages' Posts About COVID-19 Vaccination. Cyberpsychol Behav Soc Netw 2022;25:552-60.
19. Dzubur E, Khalil C, Almario CV, et al. Patient Concerns and Perceptions Regarding Biologic Therapies in Ankylosing Spondylitis: Insights From a Large-Scale Survey of Social Media Platforms. Arthritis Care Res (Hoboken) 2019;71:323-30.
20. Shapiro GK, Surian D, Dunn AG, et al. Comparing human papillomavirus vaccine concerns on Twitter: a cross-sectional study of users in Australia, Canada and the UK. BMJ Open 2017;7:e016869.
21. Idriss SZ, Kvedar JC, Watson AJ. The role of online support communities: benefits of expanded social networks to patients with psoriasis. Arch Dermatol 2009;145:46-51.
22. Hongler VNS, Navarini A, Brandt O, et al. Global trends in YouTube and Google search activity for psoriasis and atopic eczema: detecting geographic hot spots, blind spots and treatment strategies. Dermatol Ther 2020;33:e13510.
23. Reynolds KA, Pithadia DJ, Lee EB, Wu JJ. A cross-sectional study of YouTube videos about psoriasis biologics. Int J Dermatol 2019;58:e61-e2.
24. Pithadia DJ, Reynolds KA, Lee EB, Wu JJ. A cross-sectional study of YouTube videos as a source of patient information about phototherapy and excimer laser for psoriasis. J Dermatolog Treat 2020;31:707-10.