An Investigation of Asthma Experiences in Arabic Communities Through Twitter Discourse

Mohammed Alotaibi[1]*, Ahmed Omar[2]
Artificial Intelligence and Sensing Technologies (AIST) Research Center, University of Tabuk, Tabuk 71491, Saudi Arabia[1]
Department of Computer Science-Faculty of Science, Minia University, Minia, Egypt, University Street, El-Minia 1666, Egypt[2]

***Abstract*—Artificial intelligence technologies can effectively analyze the public opinions from social-media platforms like twitter. This study aims to employ the AI technology and big data to explore and discuss the common issues of asthma that patients share on Twitter platform in Arabic communities. The data was acquired using the Twitter API version 2. Latent Dirichlet Allocation was used for grouping data into two clusters which provide information and tips about the treatment and prevention of asthma and personal experiences with asthma, including symptoms, diagnosis, and the negative impact of asthma on the quality of life. Sentiment analysis and data frequency distribution techniques were used to analyze the data in both clusters. The data analysis of first indicated that individuals are interested in learning about different ways to treat asthma and potentially finding a permanent solution. The data analysis of second cluster indicated the existence of negative sentiments about asthma, which also included religious expressions for improving the condition. The study also discussed the differences in expressions among Arabic communities and other communities.**

Keywords*—Asthma; twitter; semantic analysis; LDA; Arab; communities

I. Introduction

More than 350 million people worldwide suffer from asthma, which is one of the serious public health concerns [1]. Due to changes in the environment and in people's lifestyles, its prevalence and consequences are growing in urban areas and increasing around the world. It is the most prevalent chronic childhood condition as well as one of the most expensive healthcare expenditures.

One of the most prevalent forms of reactive airway disease is asthma, which is also associated with a higher risk of death and permanent impairment [1]. Atopic dermatitis and genetic predisposition mix with eosinophilic inflammation and ongoing exposure to environmental factors, particularly molds and pollution can cause progressive lung dysfunction. Also, due to greater understanding of the biology of the disease and therapeutic advancements, asthma-related mortality has decreased over the past few decades; nonetheless, more research and efforts are required to lower asthma-related death and disability. Also, it's estimated that asthma killed more than 1000 individuals worldwide [1].

The design and delivery of healthcare systems for the management and understanding of various chronic diseases like obesity [2-4], diabetes [5-8], and asthma have been accelerated by the rapid growth of technologies, smart mobile devices, robotics, and social networks in telecommunications and the internet. A new virtual world was created as a result of the technological revolution; social networks now allow users to contact with friends and other people regardless of where they are in the world (geographically, politically, or economically). Globally, over 4.7 billion people use social networks, according to Statista [9], and this number is only anticipated to grow as mobile device use and mobile social networks gain popularity.

Twitter is one of the most commonly used social networks worldwide. It currently ranks as one of the leading social networks worldwide based on active users, according to recent social media industry statistics [9]. Twitter had 347.3 million monetizable daily active users worldwide as of the fourth quarter of 2020 [9]. Registered users can read and post tweets via the update feed, as well as follow other users [9]. This huge volume of posts on twitter platform provides billions of raw data that can be used for many purposes like research and business.

Big data is a term used to describe the enormous volume of both structured and unstructured data that regularly inundates a business [10]. Social networks in general are known as the most well-liked sources of big data. For example, each tweet posted on by a Twitter account includes multitude data input [Twitter account Id, Number of followers, Number of retweets, and Number of favorites etc.], all of which could be collected for each tweet, generating a huge volume of data in short time as the stream of data increase rapidly within seconds. Recently, artificial intelligence [AI] technology that uses huge volume of raw data has become one of the useful sources of information regarding people impressions/opinions about many events such as politics, social developments, pandemic etc.

AI technologies aid in the analysis of huge data, assisting decision-makers in their commercial decisions or governments in gaining insight into the views of the populace in their nation regarding a social, political, or economic issues. Sentiment analysis is a type of contextual text mining that can identify and extract subjective knowledge from various sources of information [11]. By monitoring online conversations, it assists a business in understanding the social perceptions of its brand, product, or service. As a result, the health sector participates in this virtual society as some patients share their experiences with the diseases they have and as some doctors have social media accounts and share clinical information with the public.

Using AI technologies for analyzing twitter conversations can be observed in healthcare research in different contexts.

For instance, twitter data was analyzed in [12] to understand the Covid-19 vaccine hesitancy; and the results revealed that potential side effects and vaccine safety were identified to be the major concerns among the public. Similarly, BERT-based supervised learning approach was used for analyzing over 31 million Covid-19 related tweets for self-disclosure in [13]. The study [13] found that users intentionally self-disclose and associate with similarly disclosing users for social rewards. Similarly, by analyzing HIV-related tweets [14] and diabetes-related tweets [15], recent research indicated that twitter discussions analysis can help in understanding nuanced public opinions, beliefs, and sentiments; and therefore, the decision-makers need to proactively use Twitter and other social media for understanding public health concerns. This is evident from a study conducted in Australia in 2019 [16]. A thunderstorm asthma outbreak in Melbourne, Australia, in 2016 led to over 8,000 hospital admissions in a matter of hours, which is a typical acute illness occurrence. A strategy based on the amount of time between events was suggested in this study since the time to respond to acute disease events is limited. Out of 18 experiment combinations, the results showed that three were able to identify the thunderstorm asthma outbreak up to nine hours ahead of the time specified in the official report, and five were able to identify it before the initial news report. The results of these studies [12-18] show the significance of Twitter monitoring and discuss conversational trends and prevailing attitudes that predominate in online social networks during a health crisis. In relation to twitter analysis of Asthma, previous studies [19-21] suggested the need for extensive research on using big data the asthma's contents in different contexts. To the best of authors' knowledge, there is no study about asthma issues in Arabic communities. Therefore, this study aims to employ the AI technology and big data to explore and discuss the common issues of asthma that patients share on Twitter platform in Arabic communities.

The remainder of the paper is organized as follows. Section II includes a review of related work. Section III describes the used methodology to develop the study. Section IV illustrates the outcomes and studies the results; and Section V discusses the results achieved while Section VI summarizes findings and outlines direction for future work.

II. Related Works

Asthma is one of the most common chronic health problems that have a significant negative influence on both society and an individual's well-being [1]. To create an epidemiological framework that can depict the condition's prevalence and patients' perceptions of that condition across multiple geographies, it is essential to integrate various large-scale data sources. Moreover, the number of social media applications has substantially increased over the last decade [20]. Twitter is a critical interactive venue for research information because statistics show that more than 80% of internet users look for health information online [9]. Social media is now being used by both patients and carers for support and information. They rely on social media for information and feedback from others to get the latest news and information on medications and treatments. Some even create and join online groups to provide support to each other.

In the contemporary era, Twitter was utilized in the health sectors, for example, to track and predict the spread of influenza [23-26]. It's also used to keep track of pharmaceutical side effects and understand the well-being of military populations [27], as well as to monitor the side effects of pharmaceuticals [28, 29]. These studies indicate the importance of social media data in public health, refining the target hypothesis' query lexicon and lowering the amount of noise in the extracted data. Despite the potential benefits, it is believed the following challenges explain why prior social media sensing experiments in public health have been short-lived or limited in scope. For example, [24] and [25] track influenza throughout a one- and two-month period, respectively. Moreover, the study in [27] investigates the harmful effects of medication over a six-month period. In terms of geographical coverage, just a few cities are examined in [24], and the transmission of influenza is studied at the national level rather than at the state or county level in [25]. Moreover, a study developed in 2013 [19] aimed to present Natural Language Processing-based Content Analysis research to aid with Asthma syndromic surveillance on Twitter. They used the Twitter API to get a big number of Tweets. Asthma and various misspellings of that word were among the search results, as were phrases for common medical devices linked with Asthma, such as "inhaler" and "nebulizer," as well as names of prescription medicines used to treat the illness, such as "albuterol" and "Singulair". Annotating the content of a randomly selected subset of these Tweets [N=3511] was done using an annotation scheme that coded for the following elements: the Asthma Symptom Experiencer [Self, Family, Friend, Named Other, Unidentified, and All-Non-Self, which was the union of these last four categories]; aspects of the type of information being conveyed by each Tweet [Medication, Triggers, Physical Activity, Contacting of a Medical Practitioner]. With the unigram model, SVM with 10-fold cross-validation achieved the highest prediction accuracy. Non-English, Self, All-Non-Self, Medication, Symptoms, and Spam were the categories with the highest reduction in classification error when utilizing the unigram model. For the unigram model, most of these categories demonstrated very high Precision and very high Recall. Surprisingly, the Unigram model performed significantly better than the bigram model, implying that individual words in these Tweets were more reliably predictive of content than pairs of words, which were less common. Authors concluded that using social media, such as Twitter, to undertake surveillance for chronic illnesses like Asthma is a promising method.

Moreover, another recent study [20] looked at the digital footprints [or "sociomes"] of asthma stakeholders on Twitter to see how they communicated online. Symplur Signals were used to collect tweets containing the word "asthma" and the hashtag #asthma. The characteristics of usage and tweets were examined between the words "asthma" and the hashtag #asthma, and then between four stakeholder groups: clinicians, patients, healthcare organizations, and industry. Authors found that with fewer people and tweets each month, the #asthma sociome was substantially smaller than the "asthma" sociome. The #asthma sociome, on the other hand, had a better correlation with asthma seasons and was less vulnerable to profanity and viral memes. Consequently, between April 2015

and November 2018, 308,370 individuals tweeted 695,980 times for the #asthma sociome. Clinicians accounted for 16% of tweets, patients for 9%, healthcare organizations for 22%, and industry for 0.3 percent. However, authors recommended that further research could aid in improving health-care communication and guiding patient and provide education.

In a different context, a recent study [21] focused on analyzing the most popular tweets and the quality of the links posted, and to determine what factors influence the debate about asthma on Twitter. The authors used Symplur Signals to extract data from Twitter, analyzing the top 100 most shared tweets and the top 50 most shared links with the hashtag #asthma. Each website's content was evaluated using an Asthma Content score, as well as validated DISCERN ratings and HONCode standards. They found out that the top 100 asthma-related tweets received 16,044 likes and were shared 10,169 times. Non-healthcare individuals accounted for 20 of the top 100 tweets, non-healthcare organizations accounted for 16, and doctors accounted for 14. There were 62 educational tweets among the top 100, 11 research-related tweets, ten political tweets, and 15 promotional tweets among the top 100. Moreover, the top 50 links were shared a total of 6009 times [median number of shares 92 (range 60-710). The most prevalent type of link was found to be instructional content (42%), followed by research papers (24%), promotional websites (22%), and political websites (12%). The Asthma Content ratings of educational links were higher than those of other links ($p=0.005$, $p< .05$). For all sorts of linkages, all three scores were poor. Only 34% of sites passed the HONCode criteria, and only 14% were found to be of good quality by the DISCERN score. The authors concluded that majority of tweets with the hashtag #asthma was educational. However, most top Twitter links rated low in terms of asthma content, quality, and trustworthiness.

A recent study [30] the impact of socio-cognitive factors on adherence to asthma medication using traditional mixed methods (interviews and twitter content analysis) and machine learning, found that some perceptions are more freely expressed on social media such as Twitter, than in the laboratory setting. Therefore, twitter data may be more reliable for understanding of public perceptions of asthma and its relevant factors compared to laboratory/hospital data in few instances. It should be noted that all studies refer to main role of tweets contents on understanding more about asthma while some studies recommend to do more search about the tweets contents towards asthma. However, this study is an attempt to contribute in adding a value to the understanding of tweets contents about asthma in Arabic communities.

III. METHODS

The study method occurred in the following phases, as shown in Fig. 1: data collection, data preprocessing (cleaning), sentiment analysis, and frequency distribution.

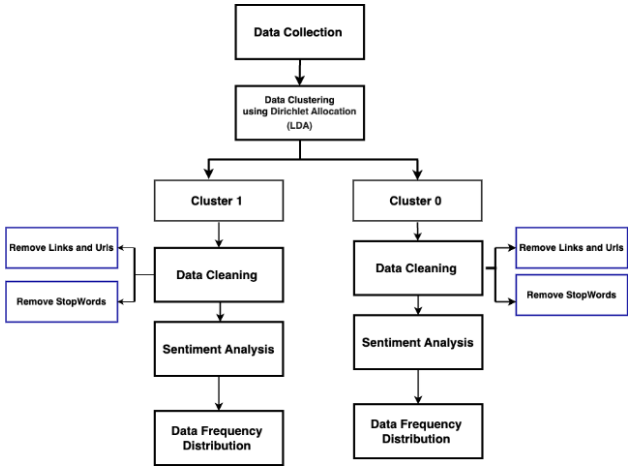

Fig. 1. Study methodology flowchart.

A. Data Collections

The data was acquired using the Twitter API version 2, premium version, which offers several additional features above the regular API version. The premium version of the Twitter API allows users to collect data from the previous 30 days. However, to enable the collection of data in excel sheet format, a python script was created. As search keywords, two separate terms [Asthma, asthmatic] were utilized. These hashtags were picked because they are popular on Twitter. User ID, user location, tweets, account followers, favorites, and retweets are all collected and kept in an excel sheet for further statistical research. One hundred thirty thousand (130,000) tweets including words asthma or asthmatic have been collected.

B. Data Clustering using Latent Dirichlet Allocation [LDA]

A statistical modeling technique called topic modeling can be used to identify the general "themes" that appear in a group of texts. A topic model such as Latent Dirichlet Allocation [LDA] is used to categorize text in a document to a certain topic. It creates a topic per document model and a words per topic model using Dirichlet distributions as the modeling framework.

C. Data Processing and Cleaning

In order to make the data clear, the following steps are followed: (1) personal interview, authors review all tweets and remove tweets that have no relation to asthma. (2) Python scripts were used to remove all tweets that include links or URLs because some of those tweets refer to another reference or for advertisement of a product etc. (3) another Python script was used to remove the stopwords in Arabic language from the tweets' contents. (4) a python script was used to tokenize tweets. Tokenization is one of the most fundamental yet crucial procedures in text analysis. Tokenization divides a stream of text into smaller pieces called tokens, which are frequently words or sentences. While this is a well-known issue with various ready-to-use solutions from popular libraries, Twitter data presents significant issues due to the language's nature.

E. Sentiment Analysis

One of the most beneficial applications of natural language processing is sentiment analysis (SA). We used "Mazajak" which is an Arabic SA system on the internet. The system is built on a deep learning model that produces cutting-edge results on a variety of datasets for Arabic dialects, including SemEval 2017 and ASTD. The existence of such a system ought to be helpful for numerous applications and fields of study that use sentiment analysis as a tool [31].

F. Data Frequency Distribution.

As it is known in every language, some words are widespread. Notably, their use in the language is crucial; they don't usually convey a particular meaning, especially if taken out of context. Therefore, in this case of data frequency distribution, stop words were removed from each tweet using python scripts; also, removing the URL was performed.

G. Anonymity and Privacy

The data (preferred as tweets) utilized in this research is freely available on the internet. However, we decided to respect the privacy of the tweet senders. As a result, the User ID of all records were removed. Perceptions were defined as socio-cognitive elements such as opinions, beliefs, and feelings in this study, and this was also the definition of perceptions employed.

IV. Results

We used LDA topic modeling to group the collected tweets into two clusters. The first cluster [cluster 0] contains tweets that provide information and tips about the treatment and prevention of asthma, including natural remedies, inhalation therapy, and the use of specific products. In contrast, the second cluster [cluster 1] contains tweets that discuss personal experiences with asthma, including symptoms, diagnosis, and the negative impact of asthma on the quality of life. There are also some tweets in this cluster that express frustration and negative feelings about asthma. In the following section we will presents and display some distributions of each cluster.

A. Cluster 0

Fig. 2 shows the distribution of sentiment in the first cluster. We can see that most of the tweets are labeled as neutral [62,366], followed by negative [32,913] and positive [6,412].

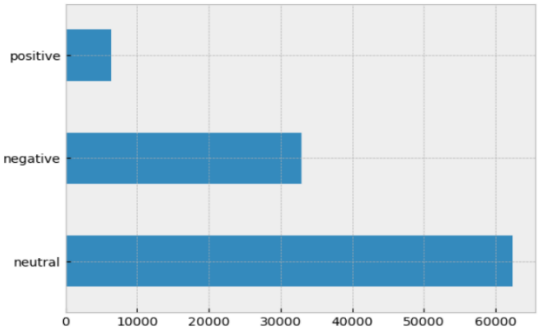

Fig. 2. Sentiment distribution in Cluster 0.

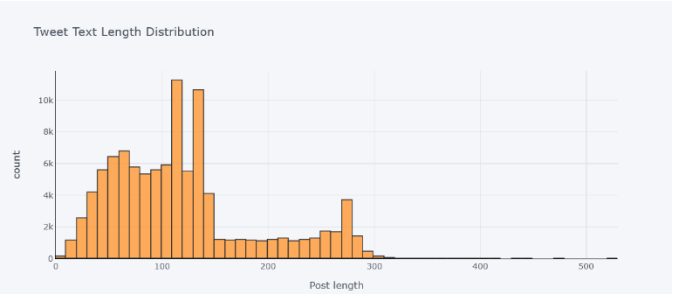

Fig. 3. Shows tweets length [number of characters] distribution.

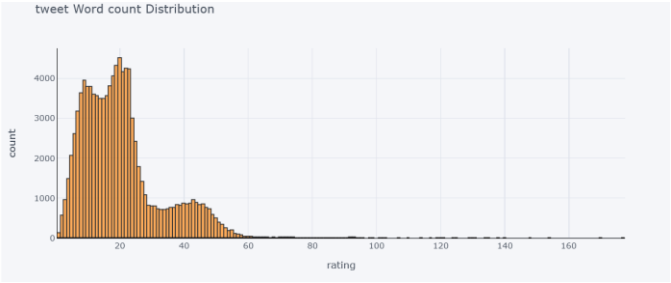

Fig. 4. Shows word count tweets distribution.

Fig. 3 shows tweets length [number of characters] distribution while Fig. 4 shows the word count tweets distribution. Top 20 words [including stop words] frequency distribution before removing stop words is shown in Fig. 5.

Fig. 5. Frequency distribution of top 20 words in Cluster 0 before removing stop words.

The data shows that the word "الربو" [asthma] has the highest frequency with 108450 occurrences, followed by "من" [from] with 52865 occurrences, and "في" [in] with 38052 occurrences.

Further analysis of the distribution reveals that the words "علاج" [treatment], "مرض " [disease], and "مرضى" [patients] are also highly frequent, which suggests that the text or corpus is likely related to medical or health topics.

It is important to note that the distribution includes some common prepositions such as "على" [on] and "مع " [with], which may not carry significant meaning on their own but contribute to the overall frequency count.

Fig. 6 represents top 20 words frequency distribution after removing stop words. In the new distribution, the word " الربو" [asthma] still has the highest frequency with 109379 occurrences, but the words "علاج " [treatment] and "مرض " [disease] have increased in frequency, suggesting that the text or corpus may be more focused on medical treatments and

conditions. Additionally, words such as "لمرضى " [for patients] and "الاطفال" [children] have been replaced with "لمرضاي " [for my patients] and "الأطفال " [kids], respectively, indicating a slight difference in phrasing.

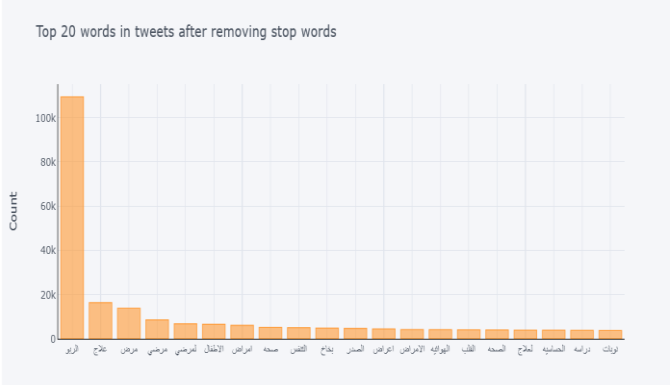

Fig. 6. Frequency distribution of top 20 words in Cluster 0 after removing stop words.

Fig. 7 shows the top 20 bigrams frequency distribution before removing stop words.

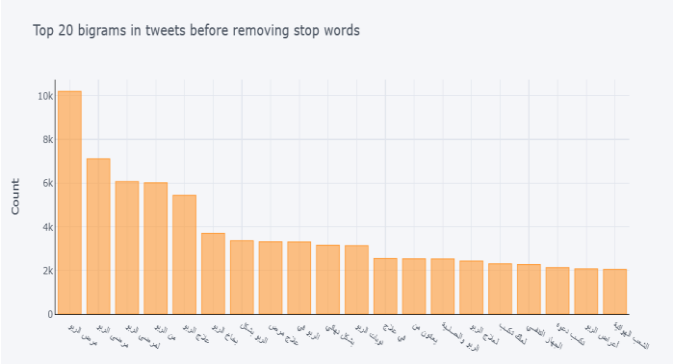

Fig. 7. Frequency distribution of top 20 bigrams in Cluster 0 before removing stop words.

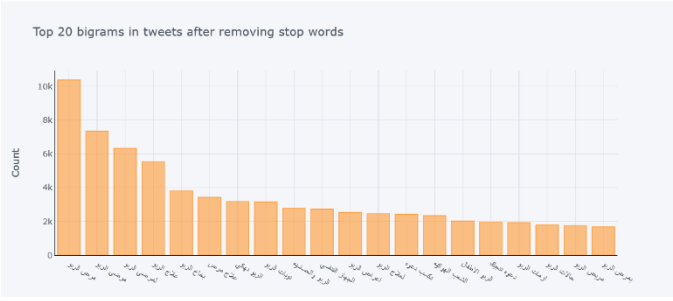

Fig. 8. Frequency distribution of top 20 bigrams in Cluster 0 after removing stop words.

The provided data is a bigram frequency distribution, which lists the frequency of two-word phrases occurring in the text or corpus. In this case, the bigrams are not filtered for stop words. The most frequent bigram is " و" [asthma disease] with 10200 occurrences, followed by "مرضى الربو " [asthma patients] with 7110 occurrences, and "لمرضى الربو " [for asthma patients] with 6067 occurrences.

When compared to the previous distribution with stop words, it is evident that the bigrams in the current distribution are more specific and related to the topic of asthma and its treatment. The bigrams also provide more context and information about the text or corpus, such as the prevalence of asthma patients and the use of inhalers as a treatment.

However, it is important to note that the inclusion of stop words in the bigrams may result in some noise and redundancy, as some common phrases that do not carry significant meaning may also appear frequently. Therefore, filtering for stop words may help to reduce noise and highlight the most meaningful bigrams which is presented in Fig. 8.

After removing stop words (Fig. 8), the bigram frequency distribution shows that "مرض الربو " [asthma disease] is still the most frequent bigram with 10366 occurrences, followed by "مرضى الربو " [asthma patients] with 7330 occurrences, and "لمرضى الربو " [for asthma patients] with 6315 occurrences.

Compared to the distribution with stop words, the current distribution has fewer occurrences of bigrams, indicating that filtering for stop words has removed noise and redundancy. The bigrams in the current distribution are more specific and related to asthma and its treatment, such as "علاج الربو " [asthma treatment] and "بخاخ الربو " [asthma inhaler].

Fig. 9 shows the top 20 trigrams frequency distribution before removing stop words.

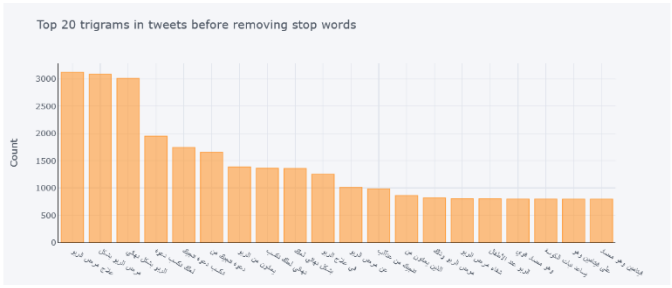

Fig. 9. Frequency distribution of top 20 trigrams in Cluster 0 before removing stop words.

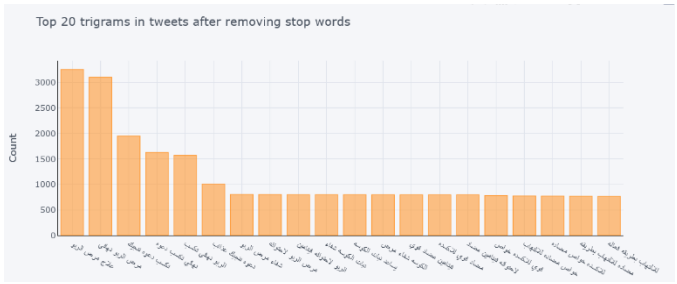

Fig. 10. Frequency distribution of top 20 triigrams in Cluster 0 after removing stop words.

The previous distribution is a frequency distribution of the top 20 trigrams related to the topic of asthma treatment. It appears that the most frequent trigrams are those related to the treatment of asthma, with "علاج مرض الربو " [treatment of asthma] being the most frequent trigram, followed closely by "مرض الربو بشكل " [asthma in a way] and "الربو بشكل نهائي " [asthma permanently]. This indicates that individuals are interested in learning about different ways to treat asthma and

potentially finding a permanent solution. Other trigrams in the list include "في علاج الربو " [in the treatment of asthma], "عن مرض الربو " [about asthma], and "شفاء مرض الربو " [cure of asthma]. These trigrams suggest that people are looking for information on various aspects of asthma treatment, including the effectiveness of different treatments, information on the condition itself, and potential cures.

Furthermore, the frequency of "يعانون من الربو " [suffer from asthma] indicates that many individuals are affected by this condition and are actively seeking ways to manage or treat it. Overall, the distribution provides insights into what people are interested in learning about regarding asthma treatment, with a particular emphasis on finding effective treatments and potentially a cure. In contrast, the trigram distribution focuses more on treatments for asthma, with "علاج مرض الربو " [treatment of asthma] being the most frequent trigram, and "علاج الربو " [asthma treatment] and "لعلاج مرض الربو " [for asthma treatment] also appearing in the list.

The distribution of the top 20 trigrams after removing stop words (Fig. 10) is different from the one without removing stop words. In this distribution, the trigrams related to asthma treatment are still present, with "علاج مرض الربو " [treatment of asthma] being the most frequent trigram. However, "مرض الربو بشكل " [asthma in a way] and "مرض الربو بشكل " [asthma permanently] are replaced by "الربو نهائيا " [asthma final] and "الربو لاحتوائه " [asthma for containing]. This suggests that people are interested in learning about the final stage of asthma and its contents.

The trigrams related to "تكسب دعوة تنجيك " [winning a prayer saves you] and "تكسب دعوة نهائي " [final, you win a prayer] indicate that most of the target population are believers and they are looking for a prayer that God [Allah] will help them and be cures from asthma. The trigrams related to "تساعد نبتة الكوسة " [helps zucchini plant] and "الكوسة شفاء مرض " [zucchini is a cure for asthma] suggest that people may be looking for natural remedies or alternative forms of treatment for asthma. The trigrams related to "مضاد قوي للأكسدة " [powerful antioxidant] and "مضاد خاص للالتهاب " [anti-inflammatory properties] indicate that people may be interested in learning about the potential benefits of antioxidants and anti-inflammatory substances in managing or treating asthma.

Overall, the distribution after removing stop words provides a different perspective on what people are interested in learning about regarding asthma treatment. While the focus on finding effective treatments and potentially a cure remains, there is also interest in the final stage of asthma, natural remedies, and potential benefits of antioxidants and anti-inflammatory substances. The word cloud for the cluster 0 is presented in Fig. 11.

Fig. 11. Word cloud for Cluster 0.

B. Cluster 1

Fig. 12 shows the distribution of sentiment in the second cluster. We can see that a majority of negative sentiment [68945] followed by neutral sentiment [23815], and a minority of positive sentiment [19013]. Compared to the first cluster, this cluster has a significantly higher proportion of negative sentiment, while the proportion of positive sentiment is also higher than the previous clusters. The majority of the sentiment being negative suggests that the text in this cluster contains a lot of negative or critical opinions, about personal experiences with asthma.

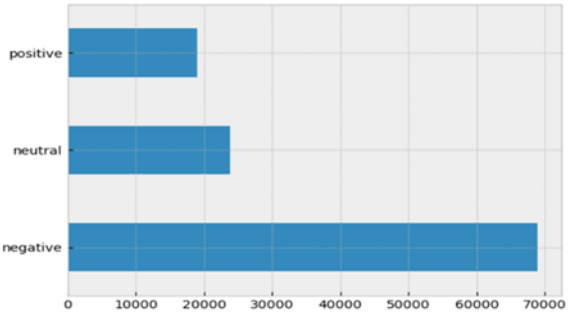

Fig. 12. Sentiment distribution in Cluster 1.

Fig. 13 shows tweets length [number of characters] distribution.

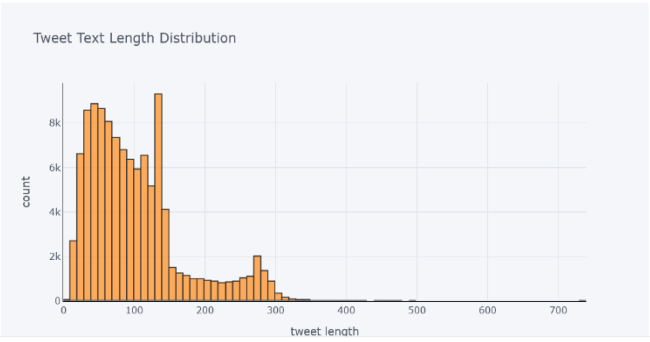

Fig. 13. Tweets length in Cluster 1.

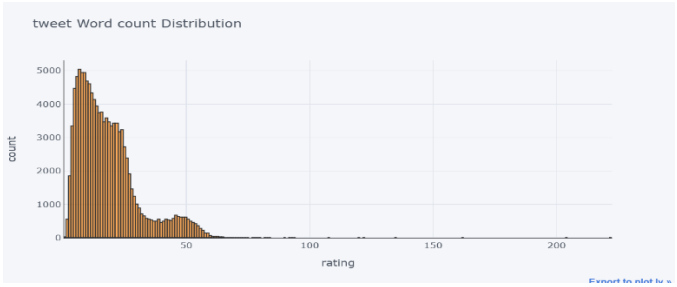

Fig. 14. Tweets distribution in Cluster 1.

Fig. 14 depicts the tweets distribution in cluster 1. Top 20 words [including stop words] frequency distribution before removing stop words is shown in Fig. 15.

Fig. 15. Frequency distribution of top 20 words in Cluster 1 before removing stop words.

The data shows that the most frequent word in the text is "الربو " [asthma] with a count of 112579, followed by The words "من " [from], "الله " [God], "في " [in], and "على " [on] with 52701, 27797, 21798, 11715 occurrences respectively. The word "ما " [what] appears in the list with a count of 11285, which suggests that the text may contain questions or inquiries related to asthma.

In the new distribution, the word "الربو " [asthma] still has the highest frequency with 109379 occurrences, but the words "علاج " [treatment] and "مرض " [disease] have increased in frequency, suggesting that the text or corpus may be more focused on medical treatments and conditions. Additionally, words such as "لمرضى " [for patients] and "الأطفال" [children] have been replaced with "لمرضاي " [for my patients] and "الاطفال " [kids], respectively, indicating a slight difference in phrasing.

Fig. 16 shows the top 20 words frequency distribution after removing stop words. The most frequent word in the text is still "الربو " [asthma], The word "لمرضي " [my illness] appears in the list with a count of 9408, which suggests that the tweets may include personal experiences of people with asthma. The word "الغبار " [dust] appears in the list with a count of 5667, which confirms that the text may be discussing asthma triggers, including environmental triggers like dust. The words "يا رب " [Oh God] and "اللهم " [O Allah] appear in the list with counts of 4724 and 4356, respectively, which suggests that some of the tweets may contain expressions of religious faith or appeals to a higher power for help with asthma management.

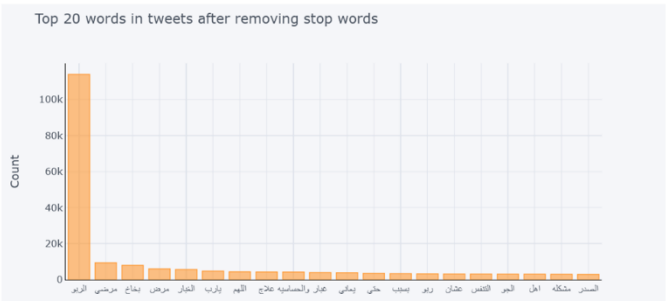

Fig. 16. Frequency distribution of top 20 words in Cluster 1 after removing stop words.

Fig. 17 shows the top 20 bigrams frequency distribution before removing stop words.

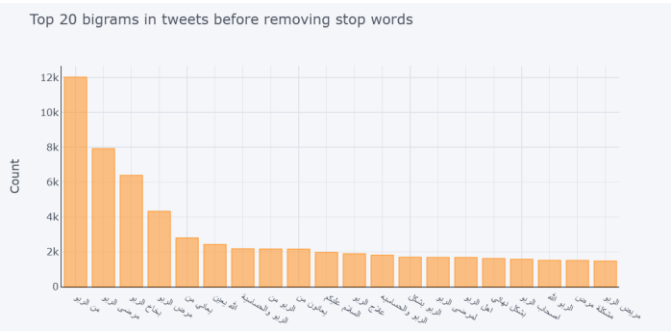

Fig. 17. Frequency distribution of top 20 bigrams in Cluster 1 before removing stop words.

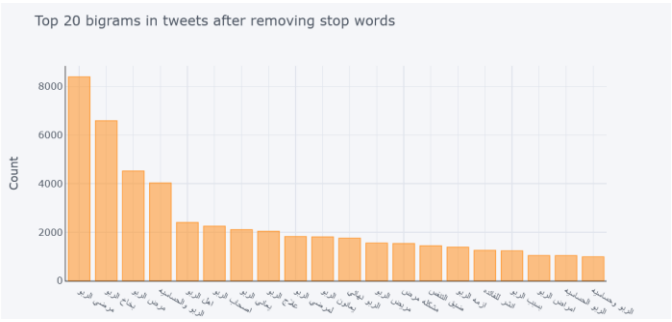

Fig. 18. Frequency distribution of top 20 bigrams in Cluster 1 after removing stop words.

The most frequent bigram is "من الربو " [because of asthma] with a count of 12014, followed by the bigram "مرضى الربو " [asthmatic patients] that appears in the list with a count of 7919, which confirms that the text is discussing asthma and may contain personal experiences of people with asthma. The bigram "يعاني من " [suffers from] appears in the list with a count of 2803, which suggests that the tweets may include discussions about the challenges and difficulties of living with asthma. The bigram "يعين الله " [God help] appears in the list with a count of 2428, which suggests that some of the tweets may contain expressions of religious faith or appeals to a higher power for help with asthma management. filtering for stop words may help to reduce noise and highlight the most meaningful bigrams which is presented in Fig. 18.

After removing stop words (Fig. 18), looking at the distribution, we can see that the most frequent bigram is "مرضي

الربو " ["my asthma" in English] with a frequency of 8394, followed by "بخاخ الربو " ["asthma inhaler"] with a frequency of 6590. These two bigrams are related to managing the symptoms of asthma and suggest that people are sharing their personal experiences with using inhalers to control their symptoms. The third most frequent bigram is "مرض الربو " ["asthma disease"] with a frequency of 4521, followed by "الربو و الحساسية " ["asthma and allergy"] with a frequency of 4029. These bigrams suggest that people are sharing their personal experiences with the diagnosis of asthma and its relationship to allergies.

Other common bigrams in the distribution include "يعاني الربو " ["suffers from asthma"], "ضيق التنفس " ["shortness of breath"], and "مشكلة مرض " ["disease problem"]. These bigrams indicate that people are sharing their personal experiences with the negative impact of asthma on their quality of life and the challenges they face in managing their symptoms.

Overall, the distribution indicates that people are sharing their personal experiences with asthma, including symptoms, diagnosis, and the negative impact of the disease on their lives. This information can be useful for healthcare providers and researchers in understanding the lived experiences of people with asthma and developing interventions to improve their quality of life.

Fig. 19 shows the top 20 trigrams frequency distribution before removing stop words.

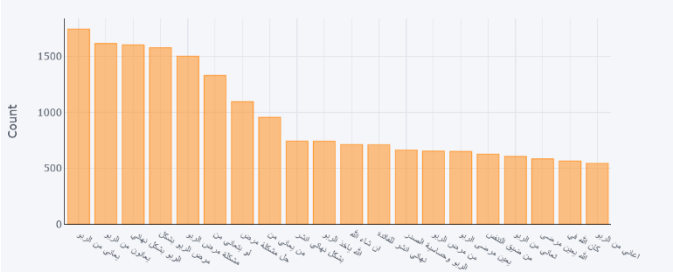

Fig. 19. Frequency distribution of top 20 trigrams in Cluster 1 before removing stop words.

Fig. 20. Frequency distribution of top 20 triigrams in Cluster 1 after removing stop words.

Looking at the trigram distribution, we can see that the most frequent trigram is "يعاين من الربو " ["suffers from asthma"] with a frequency of 1746, followed closely by "يعانون من الربو " ["they suffer from asthma"] with a frequency of 1616. These trigrams suggest that people are sharing their personal experiences with asthma and the challenges they face in managing their symptoms.

The third most frequent trigram is "الربو بشكل نهائي " ["asthma finally"] with a frequency of 1603, followed by "مرض الربو بشكل " ["asthma disease in the form of"] with a frequency of 1579. These trigrams suggest that people are discussing the long-term impact of asthma on their lives and the challenges they face in managing the disease.

Other common trigrams in the distribution include "مشكلة مرض الربو " ["asthma disease problem"], "لو تعاني من " ["if you suffer from"], and "حل مشكلة المرض " ["solve the problem of disease"]. These trigrams suggest that people are sharing their personal experiences with the negative impact of asthma on their quality of life and seeking solutions to manage the disease.

Interestingly, the trigram"الله يأخذ الربو " ["God takes away asthma"] appears in the distribution with a frequency of 741. This trigram reflects a religious or cultural belief that asthma can be cured through divine intervention.

After removing the stop words (Fig. 20), the trigram distribution seems to be more focused on specific topics related to asthma. The most common trigrams are "مرض الربو نهائيا " [asthma is final], "مشكلة مرض الربو " [the problem of asthma], and "الربو و حساسية الصدر " [asthma and chest allergy]. These trigrams indicate that this cluster is more focused on discussing the negative impact of asthma on patients' lives and the difficulties associated with managing the disease.

The trigram "بخاخ الربو يفطر " [asthma inhaler breaks the fast] appears in the distribution, indicating that this cluster includes discussions related to religious practices during the month of Ramadan. This suggests that this cluster may include personal experiences and discussions from individuals living in Islamic countries where Ramadan is observed.

Overall, the trigram distribution after removing stop words indicates that the cluster focuses on discussing the negative impact of asthma on patients' lives, including the challenges associated with managing the disease and the impact on religious practices during the month of Ramadan. The word cloud for the fists cluster in viewed in the Fig. 21.

Fig. 21. Word cloud for Cluster 1.

It is clear that this cluster focuses on personal experiences with asthma, including symptoms, diagnosis, and the negative impact of asthma on the quality of life.

V. DISCUSSION

LDA topic modeling was utilized to group tweets related to asthma into two clusters. The first cluster contained tweets that provided information and tips about the treatment and prevention of asthma. The second cluster contained tweets that discussed personal experiences with asthma and the negative impact on quality of life. Further analysis revealed that the text or corpus is related to medical or health topics, with the most frequent word being "asthma." Filtering stop words resulted in more specific and related bigrams and trigrams to asthma and its treatment. The data analysis of cluster 0 indicates that individuals are interested in learning about different ways to treat asthma and potentially finding a permanent solution. It is evident that most used phrases referred to the information on asthma, its treatments for patients and kids; with a focus on natural therapy and inhalation therapy. Similar results can be observed from [19], where it was found that most referred tweets reflected inhalation, use of nebulizer, and self-medication/ management procedures, indicating the informational content. In [20,21], it was identified that most of the tweets belonged to physicians and healthcare organizations presenting the educational and awareness information. Therefore, in similar other studies [19,20,21], the analysis of data from cluster 0 indicated that the social media platforms like twitter could be a useful platform for disseminating health information for creating awareness about asthma treatment and prevention practices, especially self-management procedures. Overall, the findings from cluster 0 could provide insights for healthcare professionals and researchers to develop better strategies and interventions for creating awareness in order to manage asthma.

In regard to cluster 1, the analysis of the sentiment, length, word count, and n-gram frequency distributions of the tweets related to asthma reveals important insights into the experiences and perceptions of people with asthma. Most of the sentiment in the second cluster is negative, indicating that this cluster contains a lot of critical opinions about personal experiences with asthma. The top words and bigrams suggest that people are sharing their personal experiences with asthma symptoms, diagnosis, and the negative impact of the disease on their lives. Filtering out stop words helps to identify the most meaningful bigrams and trigrams related to managing asthma symptoms and personal experiences of people with asthma. The analysis also highlights the prevalence of religious expressions by referring to God in the tweets related to asthma.

These findings indicate that people openly express negative sentiments about asthma and place significance importance on religion, indicating the impact of socio-cultural and religious factors among Arabic communities. However, analyzing the tweets in similar studies but in geographically different locations in previous studies [12-16,30], there were no references to the religion or god in asthma related tweets. Therefore, it is important to consider cultures in using the tweets for analyzing public perceptions related to healthcare services and disease management in order to formulate effective strategies for managing various conditions. In addition, analyzing twitter data can also be useful for assessing the public opinions related to the treatments, as in [12] vaccine hesitancy was highlighted for Covid-19. Similarly, the reactions to asthma treatment and prevention procedures can be assessed from tweets analysis among the public in order to effectively manage the condition. In [30], it was observed that public can more freely express their opinions on social media platforms than on DHP's in relation to their health conditions. Furthermore, in [13], it was observed that public express their opinions on social media to gain social rewards. This is evident from the results from cluster 1 analysis, where people in Arabic communities openly expressed their religious references and beliefs; and also, the negative impacts on their quality of life. These openly expressed views can be an important source of information for healthcare decision-makers and governments during health crisis, where an outbreak can be effectively monitored, tracked, and controlled within the time as suggested in [16]. Furthermore, the twitter data analysis can also be effectively used in other critical conditions by analyzing disease specific tweets [14,15]. Furthermore, the studies [13-25,30] discussed in this article reflected varying results at different geographical locations, while few results are similar and few contrasted with the findings in this study. Therefore, the public perceptions related to asthma challenges, its management, treatment and prevention practices may differ across the regions, and it is also to be highlighted that there could be a cultural impact on these factors as observed in this study. Therefore, it is necessary for research practitioners to frequently analyze public perceptions about asthma at regular intervals at different locations to better manage the disease. Overall, the findings from cluster 1 analysis can be useful for healthcare providers and researchers in understanding the lived experiences of people with asthma and developing interventions to improve their quality of life.

VI. CONCLUSION

This study has addressed the research gaps by discussing the public opinions of asthma in Arabic communities, thus contributing to the knowledge, which can have various practical and theoretical implications. These findings can support healthcare decision-makers in Arabic communities to better understand the asthma patients' opinions about their conditions and aid them in formulating patient-centered strategies for managing asthma. Furthermore, this study acts as a foundation or reference for future researchers in using AI technologies for analyzing public health data, especially in Arabic communities.

In conclusion, it can be observed that neutral sentiment existed in relation to asthma related information, its prevention and treatment; and negative sentiments existed on its impact on the quality of life among the Arabic communities. Although, religious/cultural influence existed in expressing the opinions and managing the conditions, it is also observed that twitter could be an effective platform for not only monitoring and controlling the disease but also to educate and create awareness among the asthma patients.

ACKNOWLEDGMENT

The authors extend their appreciation to the Deputyship for Research & Innovation, Ministry of Education in Saudi Arabia for funding this research work through the project number (S-1442-0049).

REFERENCES

[1] EAACI Global Atlas of Asthma [Internet].; 2021 [updated Accessed: 18/10/22;]. Available from: https://www.eaaci.org/newsfeed/4790-globalatlasofasthma.

[2] Alloghani M, Hussain A, Al-Jumeily D, Fergus P, Abuelma'Atti O, Hamden H. A mobile health monitoring application for obesity management and control using the internet-of-things. 2016 Sixth International Conference on Digital Information Processing and Communications [ICDIPC]; IEEE; 2016.

[3] O'Malley G, Dowdall G, Burls A, Perry IJ, Curran N. Exploring the usability of a mobile app for adolescent obesity management. JMIR mHealth and uHealth. 2014;2[2]:e3262.

[4] Wang Y, Min J, Khuri J, Xue H, Xie B, Kaminsky LA, et al. Effectiveness of mobile health interventions on diabetes and obesity treatment and management: systematic review of systematic reviews. JMIR mHealth and uHealth. 2020;8[4]:e15400.

[5] Chavez S, Fedele D, Guo Y, Bernier A, Smith M, Warnick J, et al. Mobile apps for the management of diabetes. Diabetes Care. 2017;40[10]:e145-6.

[6] Quinn CC, Clough SS, Minor JM, Lender D, Okafor MC, Gruber-Baldini A. WellDoc™ mobile diabetes management randomized controlled trial: change in clinical and behavioral outcomes and patient and physician satisfaction. Diabetes technology & therapeutics. 2008;10[3]:160-8.

[7] Rodríguez AQ, Wägner AM. Mobile phone applications for diabetes management: A systematic review. Endocrinologia, diabetes y nutricion. 2019;66[5]:330-7.

[8] Alotaibi MM, Istepanian R, Philip N. A mobile diabetes management and educational system for type-2 diabetics in Saudi Arabia [SAED]. Mhealth. 2016;2.

[9] Number of social network users of select social media platforms worldwide in 2019 and 2023Most popular social networks worldwide as of October 2021, ranked by number of active users [Internet].; 20231 [Accessed: 28/3/23updated October 10,;]. Available from: https://www.statista.com/statistics/1109866/number-social-media-users-worldwide-select-platforms/https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/.

[10] Buyya R, Calheiros RN, Dastjerdi AV. Big data: principles and paradigms. Morgan Kaufmann; 2016.

[11] Pozzi F, Fersini E, Messina E, Liu B. Sentiment analysis in social networks. Morgan Kaufmann; 2016.

[12] Malova E. Understanding online conversations about COVID-19 vaccine on Twitter: Vaccine hesitancy amid the Public Health Crisis. Communication Research Reports. 2021;38(5):346–56.

[13] Umar P, Akiti C, Squicciarini A, Rajtmajer S. Self-disclosure on Twitter during the COVID-19 pandemic: A network perspective. Machine Learning and Knowledge Discovery in Databases Applied Data Science Track. 2021;:271–86.

[14] Malik A, Antonino A, Khan ML, Nieminen M. Characterizing HIV discussions and engagement on Twitter. Health and Technology. 2021;11(6):1237–45.

[15] Akhila AM, Gayathri C, Srinivas B, Devi BSK. A review on sentiment analysis of Twitter data for diabetes classification and prediction. 2022 Seventh International Conference on Parallel, Distributed and Grid Computing (PDGC). 2022;

[16] Joshi A, Sparks R, McHugh J, Karimi S, Paris C, MacIntyre CR. Harnessing Tweets for Early Detection of an Acute Disease Event. Epidemiology. 2020;31(1):90-97. doi:10.1097/EDE.0000000000001133

[17] Ainley E, Witwicki C, Tallett A, Graham C. Using Twitter comments to understand people's experiences of UK health care during the COVID-19 pandemic: Thematic and sentiment analysis. Journal of Medical Internet Research. 2021;23(10).

[18] Shah SHH, Noor S, Butt AS, Halepoto H. Twitter Research Synthesis for Health Promotion: A Bibliometric Analysis. Iran J Public Health. 2021;50(11):2283-2291. doi:10.18502/ijph.v50i11.7584.

[19] Gillingham G, Conway MA, Chapman WW, Casale MB, Pettigrew KB. # wheezing: A Content Analysis of Asthma-Related Tweets. Online Journal of Public Health Informatics. 2013;5[1].

[20] Carroll CL, Kaul V, Sala KA, Dangayach NS. Describing the digital footprints or "sociomes" of asthma for stakeholder groups on Twitter. ATS scholar. 2020;1[1]:55-66.

[21] Kaul V, Szakmany T, Peters JI, Stukus D, Sala KA, Dangayach N, et al. Quality of the discussion of asthma on twitter. Journal of Asthma. 2020:1-8.

[22] Social media - Statistics & Facts [Internet].; 2021 [updated Feb 25,;]. Available from: https://www.statista.com/topics/1164/social-networks/.

[23] Cambria E, Das D, Bandyopadhyay S, Feraco A. Affective computing and sentiment analysis. In: A practical guide to sentiment analysis. Springer; 2017. p. 1-10.

[24] Byrd K, Mansurov A, Baysal O. Mining twitter data for influenza detection and surveillance. Proceedings of the International Workshop on Software Engineering in Healthcare Systems; ; 2016.

[25] Song S, Miled ZB. Digital immunization surveillance: monitoring flu vaccination rates using online social networks. 2017 IEEE 14th International Conference on Mobile Ad Hoc and Sensor Systems [MASS]; IEEE; 2017.

[26] Culotta A. Towards detecting influenza epidemics by analyzing Twitter messages. Proceedings of the first workshop on social media analytics; ; 2010.

[27] Freifeld CC, Brownstein JS, Menone CM, Bao W, Filice R, Kass-Hout T, et al. Digital drug safety surveillance: monitoring pharmaceutical products in twitter. Drug safety. 2014;37[5]:343-50.

[28] Tutubalina E, Nikolenko S. Exploring convolutional neural networks and topic models for user profiling from drug reviews. Multimedia Tools Appl. 2018;77[4]:4791-809.

[29] Klein A, Sarker A, Rouhizadeh M, O'Connor K, Gonzalez G. Detecting personal medication intake in Twitter: an annotated corpus and baseline classification system. BioNLP 2017; ; 2017.

[30] Abu Farha, I. and Magdy, W. [2019] "Mazajak: An online Arabic sentiment analyser," Proceedings of the Fourth Arabic Natural Language Processing Workshop [Preprint]. Available at: https://doi.org/10.18653/v1/w19-4621.

[31] Ljevar, V. Exploring the impact of socio-cognitive factors on adherence to asthma medication using traditional mixed methods and machine learning. PhD thesis, University of Nottingham, 2022.