

# The Role of Twitter in Raising Users' Awareness in the Prevention of Cardiovascular Disease

Mostafa Akhavan-Safar<sup>a\*</sup>, Mohammad Mohsen Sadr<sup>b</sup>, Mohsen Yoosefi Nejad<sup>c</sup>

Department of Computer and Information Technology Engineering, Payame Noor University (PNU), Tehran, IRAN; akhavansaffar@pnu.ac.ir<sup>a</sup>, m\_sadr@pnu.ac.ir<sup>b</sup>, m\_yoosefi@pnu.ac.ir<sup>c</sup>.

## ABSTRACT


Heart disease has emerged as the foremost global cause of mortality. Enhancing awareness and understanding of this disease, along with preventive strategies, is pivotal in averting its onset and diminishing mortality rates. Today, social networks have evolved into paramount information dissemination platforms owing to their user-friendly nature and widespread adoption. Among these, Twitter stands out as a prominent source of rich data that can be leveraged for educational and awareness purposes. While existing studies have evaluated the impact of social media on increasing health-related knowledge, there is a gap in research regarding the role of Twitter in increasing cardiovascular disease awareness and prevention from different perspectives. This study employs a cross-sectional and descriptive methodology to quantitatively analyze over 50026 tweets. This study seeks to investigate how Twitter users seek and disseminate information related to cardiovascular diseases. It aims to identify the prevalent topics shared about cardiovascular diseases and analyze the content of these messages. Initially, 50026 tweets from 8,619 users were gathered over a one-month timeframe. English tweets have been selected due to the prevalence of the English language. Subsequently, the tweets were categorized and analyzed utilizing the LDA technique and the MALLET platform. Content analysis was conducted across various categories, focusing on topics, temporal trends, and geographical locations of the tweets. The results show that there was a significant relationship between the parameters extracted in the research and the most concern of users was in the field of heart diseases and prevention methods. Most user tweets (36,323 or 72.60%) contained specific information about heart disease. 9.33% related to cardiovascular issues, 2817 (5.63%) tweets were about heart attack, 2949 (5.89%) were about heart failure and 3267 (6.5%) about other cases related to heart disorders (cardiac arrest, cardiomyopathy, ischemic heart, etc.). The most concern of users in the group of heart diseases was related to the connection of topics such as cholesterol (4102 tweets (11.04%)), prevention (20348 tweets (56.01%)) and diet (1114 (3.06%)) with heart disease.

**Keywords**— Cardiovascular Diseases, Twitter, social media, Prevention, Information.

## 1. Introduction

Heart disease and stroke were the number one and number two causes of death in the world in 2019, respectively, according to the American Heart Association's report [1]. Thirty-one percent of the world's deaths are from heart disease, with 80 percent occurring in low- and middle-income countries. According to Mark Krieger, president of the American Heart Association, despite progress in reducing deaths from heart disease and stroke, the death toll is still high. Cardiovascular disease is the

most common cause of death in most countries of the world and the most important cause of disability [1]. At the beginning of the twentieth century, cardiovascular disease accounted for less than 10 percent of all deaths worldwide. At the beginning of the 21st century, cardiovascular disease accounted for almost half of all deaths in developed countries and 25% of deaths in developing countries. By 2022, cardiovascular disease is predicted to kill more than 25 million people a year. According to the World Health Organization, in 2019 the total number of deaths due to coronary artery disease increased from 1.7 million cases in 2002 to 1.11

 <http://dx.doi.org/10.22133/ijwr.2024.430706.1199>

**Citation** M. Akhavan-Safar, M. Sadr, M. Yoosefi Nejad, " The Role of Twitter in Raising Users' Awareness in the Prevention of Cardiovascular Disease ", *International Journal of Web Research*, vol.6, no.2, pp.107-117, 2023, doi: <http://dx.doi.org/10.22133/ijwr.2024.430706.1199>.

\*Corresponding Author

Article History: Received: 15 September 2023 ; Revised: 15 December 2023; Accepted: 28 December 2023.

Copyright © 2022 University of Science and Culture. Published by University of Science and Culture. This work is licensed under a Creative Commons Attribution-Noncommercial 4.0 International license(<https://creativecommons.org/licenses/by-nc/4.0/>). Noncommercial uses of the work are permitted, provided the original work is properly cited.

million cases. For this reason, the need to inform and raise public awareness in order to prevent cardiovascular disease has become particularly important. In many cases, proper knowledge and information is very effective in preventing heart disease or even saving the lives of people who have suffered a heart attack or cardiac arrest. The use of new information tools such as social networks in the present age can be useful in informing and sharing people's experiences.

Social networks are one of the most important and modern mass media [2]. One of the important consequences of the presence and participation of users in these networks is the emergence of mass movements and crises, or its control [3]. One of the most popular and famous of these networks is Twitter [4]. According to a report released by the Pew Research Center and the Knight Foundation, news following through social networks Facebook and Twitter has increased dramatically [5]. Nearly two-thirds (63%) of Twitter and Facebook users said they follow the news through their social networks. [6] For Twitter, that number has risen 11 percent since 2013, when 52 percent of its users followed the news. For Facebook, the statistic is 16 percent higher than in 2013, when 47 percent of users said they used it to follow the news. According to the latest statistics available in 2018, Twitter has more than 321 million monthly active users. In other interesting words, if Twitter declares independence as a country with the number of existing users, it will be the twelfth most populous country in the world. The number of tweets sent by these users per day is more than 500 million tweets. Due to the spread of this social media among the people and the large amount of information that is exchanged or shared through it, it can be used to analyze needs, information, increase user awareness and analyze their interests and feelings. For instance, Twitter offers networking, support, collaboration, and educational opportunities for professionals in various specialties such as cardiology, hematology, medical oncology, and radiation oncology [7]. There has been a lot of research lately on analyzing information on Twitter.

Many people are looking for useful information on topics related to cardiovascular diseases and their prevention. Most of these people are too busy to see a medical specialist. As a result, Twitter can serve as a suitable tool to disseminate valuable information about the factors influencing cardiovascular diseases, enhance user awareness, and contribute to public health management. The primary inquiries in this study are: firstly, which groups are active on Twitter within the realm of cardiovascular diseases, secondly, what topics are most prevalent in their discussions; and finally, in which geographic regions are Twitter users most active, and at what times do they tend to post more tweets. The

following is a review of the relevant literature, as well as the research gaps in this field.

## 2. Related Work

Various studies have been conducted so far regarding the use of Twitter content in increasing awareness in the field of health and other fields. For example, in the research conducted in 2020, researchers analyzed the content of tweets of users regarding weight loss and fitness [8]. The obtained results showed that there is a significant relationship between the extracted parameters regarding weight loss and fitness and the content of users' tweets. In a separate study conducted in 2020, researchers utilized Twitter content analysis to forecast childhood cancer (CCAM) [9]. The findings revealed that four distinct groups utilize social media to raise awareness about childhood cancer during CCAM. In a separate study, May et al. attempted to identify weight loss support networks within social media platforms by analyzing tweets concerning weight loss. The study involved the creation of four Twitter accounts over a span of 2 to 5 weeks, designed to follow specialized health centers and their associated user accounts [10]. Another study conducted in 2021 explored the impact of Twitter on enhancing awareness and knowledge about diabetes among the population of Saudi Arabia [11]. The results showed that Twitter had a positive effect on increasing diabetes awareness and knowledge among adult Twitter users in Saudi Arabia. In research in 2019, Twitter content analysis was used as a tool to evaluate the activities that are carried out regarding antibiotics [12]. They qualitatively determined the most influential users; the use of hashtags and the flow of information were also analyzed (the most retweeted users and the global network formed by all users). In a study conducted in 2022, the authors investigated the efficacy of Twitter in raising tax awareness during the COVID-19 pandemic [13]. The results showed that the tax payers' awareness campaign conducted by the General Tax Administration is very good in providing tax-related information and training. In another study in 2022, Twitter was studied as a tool to increase social awareness in the field of children's emergency mental health [14]. The results showed that Twitter can be an effective tool to increase research awareness in the field of PEM mental health care. Also, several success strategies were introduced in this field. In another study in 2023, the use of Twitter data to assess awareness in the field of pangolin conservation was investigated [15]. In this study, a method was introduced to analyze social media content on Twitter to determine the sentiments and identify the demographic interest of environmentalists. They used the LDA technique for topic modeling and sentiment analysis to further measure awareness and understanding around

pangolins. The results only show a significant group of people tweeting about pangolins.

In the field of cardiovascular diseases and the role of Twitter in increasing awareness and reducing mortality, research has also been done. In the study conducted by Claire et al., the role of the support campaign called " #CardioOncology" on Twitter to raise awareness about cardiovascular diseases was evaluated [16]. They analyzed quantitative and qualitative data to support the effectiveness of using Twitter for such educational campaigns and described key features that contributed to its success. Another 2019 study looked at the use of Twitter to raise awareness of familial blood cholesterol [17]. They measured FH-related Twitter activity from August to October using Sprinklr Listening Explorer software to extract tweets and engagement metrics, including reach. As a control, the reach of tweets related to colorectal cancer in the same period was also checked. The significant increase in FH-related Twitter metrics indicated that Twitter has the potential to promote FH awareness. Also, the obtained results have motivated future research on the use of social media to improve the diagnosis of FH and thus reduce morbidity and mortality. In another study, the use of Twitter as a potential data source for cardiovascular disease research was studied [18]. Their aim was to describe the volume and content of tweets related to cardiovascular diseases and the characteristics of Twitter users. The obtained results showed that the expressions of diabetes and myocardial infarction were used more than the expressions of heart failure. Most of the tweets were related to health. Common themes included tweets related to risk factors, awareness, and disease management. In 2020, another study was conducted in the field of understanding Twitter analytics in cardiology [19]. Their research showed that Twitter can be considered as a framework for dealing with the use of social media in the cardiovascular field. Also, a better understanding of Twitter analytics will be the key to the continuous growth and success of this media in the field of cardiovascular diseases.

In this study, Latent Dirichlet Allocation (LDA) technique and MALLET platform were used to analyze and categorize users' tweets about cardiovascular diseases. We do this by identifying different groups of people tweeting about cardiovascular diseases over time. While existing studies have evaluated the impact of social media on increasing health-related knowledge, there is a gap in research regarding the role of Twitter in increasing cardiovascular disease awareness and prevention from different perspectives. previous research has addressed the need for communication between physicians and cardiologists in the use of social media, including Twitter, to expand knowledge and experience [18, 20-21]. But our

study is unique in that it not only categorizes tweets related to cardiovascular diseases but also analyzes them based on user perspective, geographic location, temporal and spatial trends, and message content.

### 3. Research methodology

#### 3.1. Data gathering

This study employed a cross-sectional descriptive approach [22]. To monitor tweets concerning cardiovascular disease, a month-long collection of relevant tweets was gathered from November 4 to December 4 by utilizing specific keywords. To do this, same [23], the Twitter Application Programming Interface (API) was used. The key search terms utilized are outlined in Table 1.

Alongside the tweets, we gathered the tweet author's ID, the tweet's language, and the timestamp of when the tweet was posted. Subsequently, utilizing the author IDs obtained earlier, we employed the Twitter API to retrieve the account biographies of users who authored tweets concerning pangolins. To maintain the anonymity of the accounts, no usernames or locations were utilized in this study. In total, 50026 tweets about cardiovascular disease sent by 8619 users were collected. All tweets were composed in English language. Several steps were implemented to standardize the textual data before topic modeling, which is outlined in the preprocessing step in Figure 1. These steps are crucial for enhancing the efficiency and interpretability of the model output.

The block diagram of the research method is shown in Figure 1 shows the Shad data.

Table 1. The key search terms used to collect tweets

#cardio disease	#Cardiovascular	#CardiovascularDisease
#worldHeartDay	#Heartdiseaseday	#Cardiomyopathy
#Heartattacks	#heartFailure	#ischemicHeartDisease
#cardiacArrest	#HeartDisease	#Heartstroke

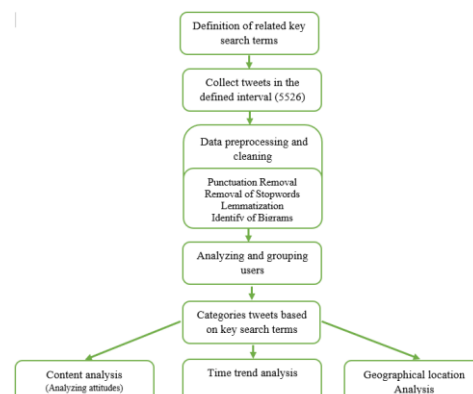


Figure. 1. Block diagram of research steps

**3.2. Statistical analysis of collected data**

The total number of extracted tweets amounted to 50,026, which were associated with 8,619 users. Twitter users were classified into organizations, individuals, and newsgroups. Organizations encompassed users with profile or screen names containing words such as foundation, association, organization, community, research, center, division, university, or hospital, with a follower count exceeding 5,000. These keywords were selected to denote organizations based on the content of Twitter profiles and usernames of users discussing heart disease. The second category comprised celebrities. Celebrities were users who did not meet the criteria for organizations but had a substantial following and/or verified tweets. Organizations constituted 10.01% of users, individuals accounted for 0.05%, newsgroups made up 3.41%, and the majority, 85.5%, belonged to the general public. The classification results indicate that the highest number of users are ordinary individuals who have posted the most tweets regarding cardiovascular diseases, highlighting the significant interest and engagement of the general public in this domain. Figure 2 illustrates the distribution of users across organizations, newsgroups, celebrities, and the general public.

On Twitter, users have the option to either compose an original tweet or retweet posts from other users. In the case of retweets, these posts typically begin with "RT@username" or include "via@username". Retweets are a distinctive form of mention where a user shares content created by others. Apart from retweeting, users can also mention other users in their tweets by using the "@username" format, where "username" refers to the Twitter handle of the mentioned user. A single tweet may be mentioned by many users. Therefore, we categorized each tweet as original or a retweet. Users in the retweet group are those who have the phrase "@username" in their descriptions, who retweeted 7096 (14.19%) and 42930 (58.81%) retweeted the original. And then we determined whether it includes a mention or not? The statistics for this category of tweets are shown separately in Figure 3.

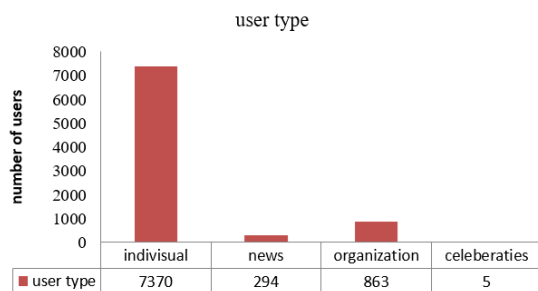


Figure 2. Categorize users by organizations, celebrities, newsgroups, and commonality

As observed, 42,930 tweets (85.81%) were main tweets, while 7,096 tweets (14.19%) were retweets. Among the retweets, 1,983 (4.61%) included a mention, and 40,947 (95.39%) did not contain any mentions. Within the main tweets, 599 (8.44%) had mentions, while 6,497 (91.55%) did not include any mentions.

**4. Categories of tweets**

Using the search terms mentioned in section 3, reviewing previous research and also consulting with several cardiologists at Mashhad University of Medical Sciences, the category of tweets was determined. Tweets were initially categorized as related or unrelated to cardiovascular disease. Unrelated tweets were categorized as miscellaneous. Tweets containing non-English words or terms were deleted.

Subsequently, the tweets were classified by topic into one of the following five categories:

1. Tweets pertaining to heart disease (heart disease)
2. Tweets concerning cardiovascular topics (cardiovascular)
3. Tweets addressing to heart attacks (heart attack)
4. Tweets related to heart failure (heart failure)
5. Tweets focusing to other heart topics such as cardiomyopathy, ischemic heart disease, cardiac arrest, ...

We also categorized and reviewed the tweets based on the following:

- Based on content
- Based on publication
- Based on time trend

We also categorized the authors' purpose of publishing a tweet based on the type of novelty they recorded as follows:

- Dissemination of personal experiences
- Information sharing
- Information Searching

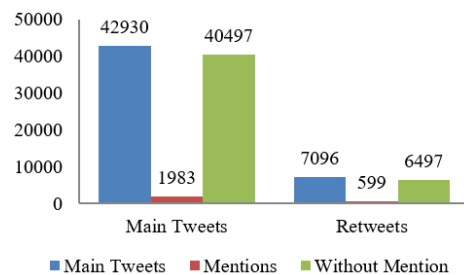


Figure 3. Number of mentions and tweets in the main tweets and retweets

The content analysis of tweets was conducted using the Machine Learning for Language Toolkit (MALLET) technique [24]. MALLET is a Java-based toolkit designed for natural statistical language processing, document classification, clustering, topic modeling, data mining, and other text-based machine learning applications. This toolkit employs the Latent Dirichlet Allocation (LDA) algorithm for pattern design and analysis, making it well-suited for handling large volumes of textual data, such as the dataset utilized in this study. Additionally, complementary tools like the Apache web server, MySQL database, and XAMPP software were employed for data preprocessing and other necessary analyses. XAMPP is an open-access web server comprising the Apache HTTP server, the MySQL database, and interpreters for PHP and Perl scripts. Shared personal information refers to messages that appear to be directly related to a person's experiences and include some related pronouns such as "I", "My", "OUR", "US" and "They".

Tweets showing the sharing of public information were categorized without focusing on one person, and these were tweets that included words like "you", "others" and "the public".

Question tweets were categorized as information search. This tweet included words such as "which", "what", "how", "why", "did you", and so on. For example:

Do you know what risk factors contribute to the development of heart disease?

Additionally, tweets falling outside these five categories underwent further review and were evaluated by three experts. Subsequently, these tweets were assigned to appropriate designated categories based on the experts' assessments.

## 5. Results

In this section, the results of various analyzes are shown.

### 5.1. Results from the classification of tweets

As previously detailed, a total of 50,026 tweets were collected using 12 specified search terms. Among these, the majority (36,323 tweets or 72.60%) contained explicit information concerning heart disease and were thus included in the final study group. Of the total tweets, 4,672 (9.33%) were associated with cardiovascular issues, 2,817 (5.63%) focused on heart attacks, 2,949 (5.89%) discussed heart failure, and 3,267 (6.5%) were dedicated to other heart-related conditions such as cardiac arrest, cardiomyopathy, and ischemic heart disease. The distribution of tweets across these categories is illustrated in Figure 4.

Also, the amount of tweets, retweets and mentions in each category was calculated, which is shown in Figure 5.

Given that the heart disease category had the highest percentage of tweets, we conducted a detailed analysis to identify specific mentions related to heart disease within this group. Among these, 4,102 tweets (11.04%) discussed cholesterol, while 20,348 tweets (56.01%) addressed prevention issues. Additionally, 1,114 tweets (3.06%) focused on diet and its correlation with heart disease. The relevant results are shown in Figure 6. 1220 (3.35%) tweets mentioned diabetes and heart disease. 7,366 (20.27%) tweets mentioned the risks associated with heart disease (such as salt intake, children, etc.). In general, among the tweets related to heart disease, 56.01% for prevention, 20.27% for risks related to heart disease and 11.04% for cholesterol and heart disease and 12.40% was for other cases related to heart disease such as smoking, salt, sugar, diet and .... Moreover, we calculated the distribution of tweets, retweets, and mentions within each category, as depicted in Figure 7.

### 5.2. Time trend analysis of tweets

We conducted an analysis of the volume of tweets posted on various days of the week, categorized by date and time. The findings from this analysis are illustrated in Figure 8.

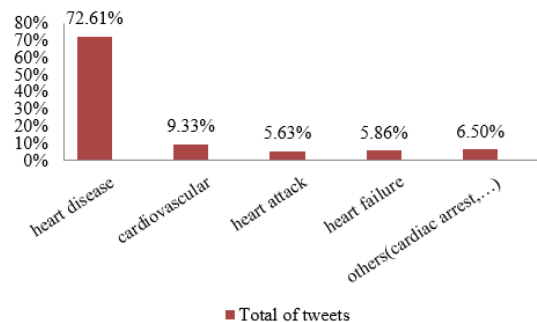


Figure 4. Classification of Tweets by Section 4 Category

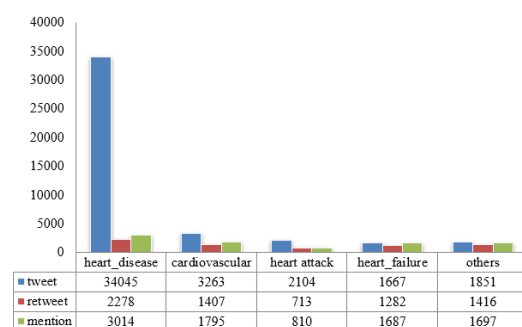


Figure 5. Number of mentions, tweets and retweets by topic category

The chart indicates that the highest number of tweets were posted on Fridays and Tuesdays, while the lowest number of tweets were observed on Wednesdays. Consequently, it appears that the majority of tweets concerning heart disease are posted during the weekends. Additionally, we extracted the total number of tweets posted over a one-month period (data collection period), as depicted in Figure 9.

The data presented in Figure 8 reveals a rising trend in the number of tweets posted from the start of December to the middle of the month, peaking on December 15. To investigate the cause of the issue, we reviewed research related to heart disease and found that research in the Journal of the American Heart Association (AHA) [14] showed that deaths from heart attacks increased during the December holiday season and winter holidays as is a risk factor for heart death. Therefore, the increase in tweets posted on the above date shows the accuracy of this issue.

**5.3. Content analysis of tweets**

In order to conduct a thorough content analysis of the tweets, we initially focused on the content of tweets within the heart disease domain, which constituted a substantial proportion (72.61% of all tweets). Among these, 318 tweets (1%) pertained to individuals sharing personal experiences, while 3,876 tweets (67%) involved users disseminating general information about heart disease. Furthermore, 1,405 tweets (3.86%) were categorized as inquiries where users posed questions about heart disease. The distribution of these tweet categories and exemplar tweets are detailed in Table 2.

Subsequently, this analysis was conducted based on the classification outlined in Figure 6. As shown in Table 3, tweets related to prevention of heart disease were first categorized.

As shown in Figure 10, the analysis revealed that out of 20343 tweets, 73 (0.35%) tweets related to personal information about heart disease prevention, 2257 (11.09%) were tweets in which users shared their general information about the prevention of heart disease. Additionally, 1,022 tweets (5.02%) comprised inquiries from users seeking information on heart disease prevention. The majority of tweets, 16,996 (83.54%), were dedicated to discussions surrounding the knowledge of heart disease prevention practices.

Out of, 7366 tweets related to the risk of heart disease, 19 (0.25%) tweets related to personal information about heart disease risks. 1658 (22.50%) included tweets in which users shared general information about the risk of heart disease. Furthermore, 1081 (14.67%) tweets were related to tweets in which users asked questions about the risks associated with heart disease. 4608 (62.55%)

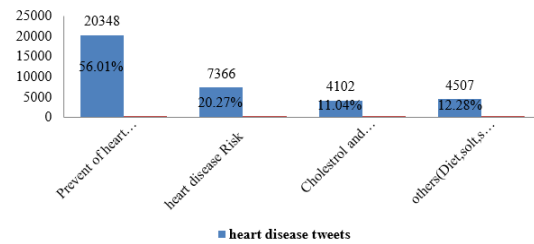


Figure. 6. Percentage of tweets related to heart disease topics

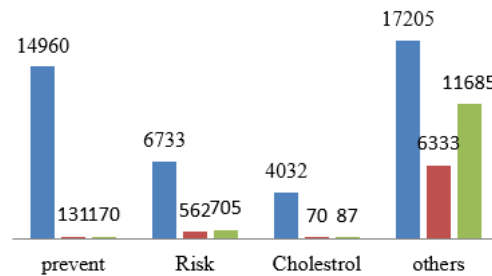


Figure. 7. Number of mentions, tweets and retweets by classification of topics related to heart disease

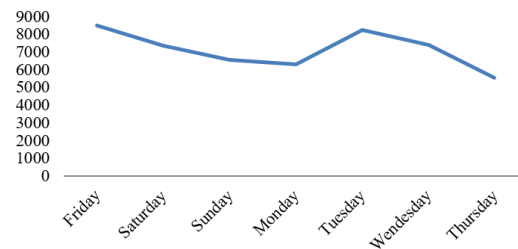


Figure. 8. The number of tweets sent on different days of the week

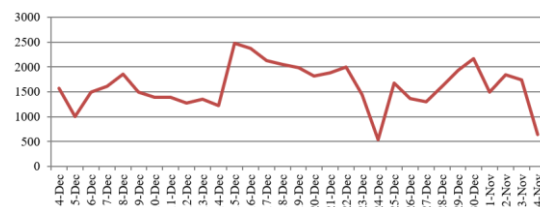


Figure. 9. The number of tweets sent on different days

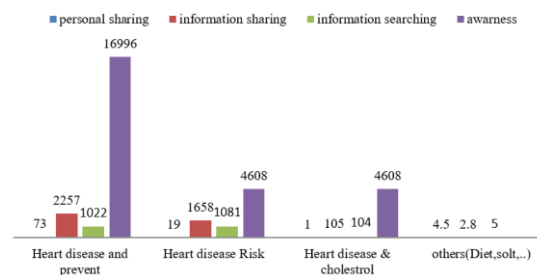


Figure. 10. Comparing the content of tweets

Table 2. Characterization of heart disease related tweets by category.

<i>category n (%)</i>	<i>Description</i>	<i>Example tweets</i>
<b>Heart disease(n=36323)</b> Personal sharing 318(1%)	Tweet shares information about a heart disease event by a person.	‘@MwamburiV: My question is what can I do in order to avoid heart disease’ ‘Not only does palm oil production destroy our rainforests and habitats for indigenous species, it can promote heart disease’ ‘Humbling to hear our patients’ stories of those affected by heart disease and stroke at AHA16.’ ‘Want to prevent heart disease & stroke? Our 3 major tips: <a href="https://t.co/b11AkYYbKt">https://t.co/b11AkYYbKt</a> <a href="https://t.co/cKFDMtgnpu">https://t.co/cKFDMtgnpu</a> ’ ‘Congenital heart disease in kids isn’t “just” about the heart’ ‘Heart Disease is the leading cause of death in the U.S. Fill out our risk calculator today and determine your risk’ -----
Information sharing 3876 (10.67%)	Tweet shares general information about a heart disease event.	‘If you have thyroid problems, you’re also at an increased risk for heart disease’ ‘The more you sit around not exercising, the more at risk you are for heart disease’ ‘@iHeartSinai: Do you suffer from heart disease? If so, you may be at risk for this complication’ ‘Did you know olive oil reduces the risk of heart disease?’ ‘5 Ways You Can Stay Heart Healthy While at Work <a href="https://t.co/KPc89Hxofo">https://t.co/KPc89Hxofo</a> ’ ‘Hypochondriacs have a higher risk of heart disease. Yes, worrying can really make you sick!’ ‘If you have anemia, you’re also at an increased risk for heart disease. More risk factors...’ ‘Women If you don’t already know the heart disease statistics, you’ll be SHOCKED’
Information searching 1405(3.86%)	Tweets have question related about heart disease	‘Do you know what risk factors contribute to the development of heart disease?’ ‘Do You Know What Will Happen If You Eat Only One Pomegranate Today?’ ‘@ADAPTABLE study: What is the optimal dose of aspirin for heart disease?’ ‘What can I do in order to avoid heart disease?’ ‘Why is heart disease still killing millions every year?’ ‘Why should you eat a plant-based diet?’ ‘What about the statins for primary prevention of heart disease?’

were related to tweets about awareness about the risk of heart disease.

Within the 4,102 tweets concerning heart disease and cholesterol, 105 tweets (2.55%) provided general information about cholesterol and heart disease, while 104 tweets (2.53%) consisted of user inquiries regarding cholesterol and heart disease. The majority of tweets, 3,893 (94.92%), were centered around raising awareness about the risks associated with heart disease.

The results of the analysis of tweets showed that among the 100 tweets that had the most repetitions, the heart attack group had the highest repetition. The chart of the most retweets according to the category of Figure 6 is shown in Figure 11.

#### 5.4. The Geographical Location Analysis Of Users

Tweets sent by users were analyzed in terms of geographical location. To identify the geographical location of users, the location and time zone fields, which have set by users on their personal page, were used. To find the location, first the location field was considered and in the absence of information, the time zone field was used. For some users, both fields were valueless, in which case the users' location was

calculated using PHP language and connected to [maps.googleapis.com](https://maps.googleapis.com), and placed in the loca(In the user profile) field. The results of the survey and analysis of users' geographical location showed that all tweets were sent from 92 countries. The most tweets were from users in the United States (30.91%), Canada (25.34%), the United Kingdom (14.92%), India (3.46%), Australia (3.18%) and Spain (2.48%). The results for the number of tweets sent from each country are shown in Figure 12.

Additionally, the percentage of tweets sent from each continent was calculated, which is shown in Figure 13. The highest number of tweets sent regarding cardiovascular diseases was related to Europe (35%) and then Asia (30%) and the lowest number of tweets was related to Oceania (2%).

The highest number of followers with 7255499 was related to a user from Ecuador, followed by a user from South Africa with 1276512. Users with more than 100,000 followers were identified and their geographical range was calculated. The number of followers and the geographical range of users with more than 100,000 followers are shown in Figure 14.

Table 3. Characteristics of heart disease-related tweets by category Figure5.

<i>Category n (%)</i>	<i>Description</i>	<i>Example tweets</i>
<b>Heart disease and prevention (n=20348)</b> Personal sharing 73(0.35%)	Tweet shares information about a heart disease event by a person.	'@TheBHF, I assume you are promoting a vegan lifestyle to prevent and reverse heart disease' 'Our 3 major tips to prevent heart disease & stroke'
Information sharing 2257 (11.09%)	Tweet shares general Information about prevent of heart disease	'These 5 foods can help you reduce stress, prevent heart disease and ward off certain cancers' 'Heart disease is one of the biggest killers – but by doing THIS every day you can prevent early death' '5 Risk Factors You Should Avoid to Help Prevent Heart Disease Death' 'High Levels of 'Good Cholesterol' may Not Prevent Heart Disease'
Information searching 1022(%5.02)	Tweets have question related prevent of heart disease	How to start healthy habits to prevent heart disease?' 'How to Prevent Heart Disease & Cancer Associated with Inherited Mutations?' 'How to prevent calcification of heart tissue?'
General information 16996(%83.54)	Tweets that awareness about prevent of heart disease	'Drug proven to prevent, reduce heart disease' 'FAST, EASY RECIPE THAT CAN PREVENT THE RISC OF HEART DISEASE'
<b>Heart disease Risk(n=7366)</b> Personal sharing 19(0.25%)	Tweet shares information about event risk in heart disease by a person.	'Fill out our risk calculator today and determine your risk'
Information sharing 1658 (22.50%)	Tweet shares Information about heart disease risk.	'Controlling risk of heart disease when you have diabetes' 'RT @HerHeartCare: 5 tips to help you reduce your risk for heart disease: <a href="https://t.co/7jTGc5gbuk">https://t.co/7jTGc5gbuk</a> ' 'The physical activity involved in gardening can help reduce your heart disease risk'
Information searching 1081(%14.67)	Tweets have question related of heart disease risks	'Do you know what risk factors contribute to the development of heart disease?' 'Did you know olive oil reduces the risk of heart disease?'
General information 4608(%62.55)	Tweets that awareness about heart disease risk	'Children at risk of heart disease, eating too much salt, study says' 'Heart disease risk may be lowered with recreational and commuter biking'
<b>Heart disease &amp; cholesterol(n=4102)</b> Personal sharing 0(0%)	Tweet shares information about event by a person (Cholesterol & heart disease).	-
Information sharing 105 (2.55%)	Tweet shares Information about cholesterol & heart disease.	'Raising 'good' cholesterol may not prevent heart disease' 'You have greater risk for #HeartDisease if you have low HDL cholesterol'
Information searching 104(%2.53)	Tweets have question related of cholesterol & heart disease	'How Well Do Cholesterol-Lowering Drugs Actually Work?' 'Are you suffering from heart disease? Cholesterol-lowering drug may help'
<b>General information 3893(%94.92)</b>	Tweets that awareness about cholesterol & heart disease.	'More good cholesterol may not prevent heart disease, but less bad cholesterol may' 'RT @mercola: Low Levels of "Good" #Cholesterol Connected to Many Health Risks, Not Just'



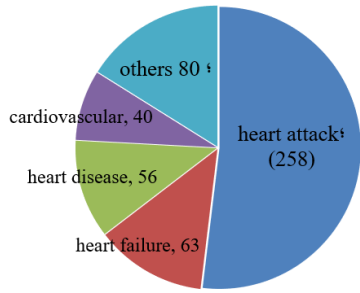


Figure 11. The number of retweets according to the classification of Figure 5.

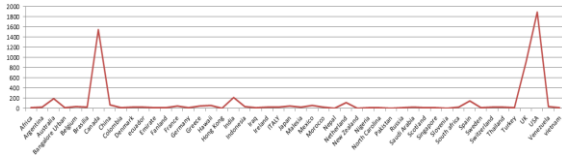


Figure 12. The number of tweets sent by each country for countries with more than 10 users

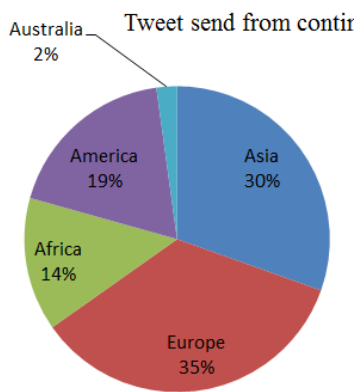


Figure 13. Amount of tweets sent by each continent

We also created a graph for users to visualize it using graphic software and did some analysis in the field of social networks. An example of graph clustering for newsgroups, organizations, and individuals after clearing a graph is shown in Figure 15. As can be seen, the highest rate is related to ordinary people.

Furthermore, the Visualizing of user network was done based on the mentions they made in order to identify nodes with high centrality and importance, which are shown in Figure 16.

As shown in Figure 15, most of the centralities belong to the American Heart Institute and a person named Dr. Martha Gulati. But the number of triangles (201 triangles) among the nodes connected to Dr. Martha Gulati is higher than the number of nodes of the American Heart Institute (18 triangles), which indicates that people connected to Dr. Martha Gulati node have more dialogue between them. Therefore, due to the high centrality and the large number of triangles, it can be considered one of the most effective nodes in

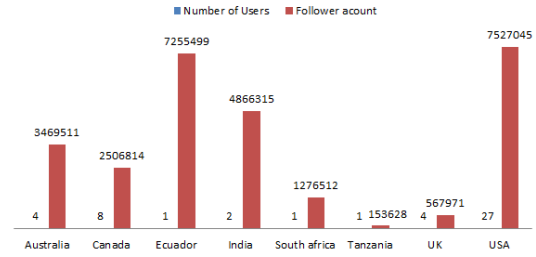


Figure 14. Number of followers and geographic range of users who had more than 100,000 followers

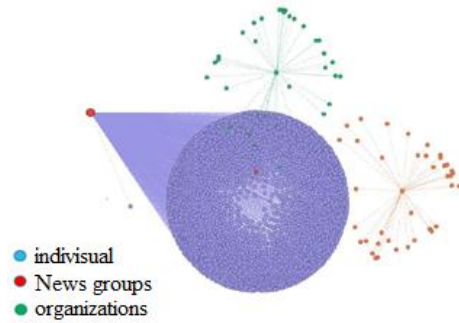


Figure 15. Visualizing and clustering of different groups of users

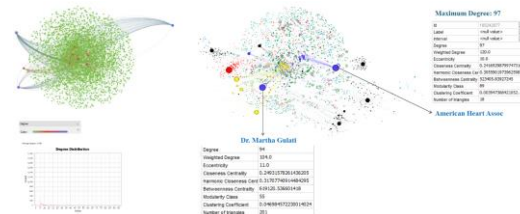


Figure 16. Network Visualizing of users who have tagged each other's posts

the discussion of information and education and increase user awareness.

## 6. Conclusion

Twitter offers excellent capabilities for various users and organizations, enabling them to educate, raise awareness, and effectively manage cardiovascular disease. In this study, we gathered tweets related to cardiovascular disease between November 4 and December 4 using specific search terms. We categorized users to determine the percentage of each group that tweeted, classified tweet content into five categories, and analyzed the content of each category based on tweet type. The results highlighted users' primary concerns within each category and the most frequently tweeted topics. Moreover, we conducted a time trend analysis of the tweets. The findings revealed a peak in mid-December, aligning with research from the American Heart Association indicating an increase in heart disease-related deaths during the December holidays. The analysis of tweets and user activity underscores Twitter's potential as a highly effective

medium for reducing heart disease and mortality. These study results can assist organizations in leveraging social media more effectively to enhance user awareness and prevent heart disease and related fatalities

## 7. Future Work

Utilizing the findings of this study, numerous avenues for future research can be pursued. One such avenue involves conducting sentiment analysis on tweets, complemented by the classifications established within this study. Furthermore, examining users' sentiments based on their geographic locations can illuminate societal attitudes towards targeted prevention of cardiovascular diseases, shedding light on variations influenced by nationality, location, gender, or profession. Another potential area for future research involves analyzing subjectivity in tweets, specifically the balance between personal opinion and factual information. Gaining insight into the evolving ratio of factual content to opinion-based content can aid in addressing misinformation within this domain. Finally, this approach can be used to inform and prevent other dangerous diseases and the results can be compared with each other. Should the same results be replicated, a focused framework could be formulated for developing campaigns aimed at enhancing awareness and promoting disease prevention through Twitter.

## Declarations

### Funding

This research did not receive any grant from funding agencies in the public, commercial, or non-profit sectors.

### Authors' contributions

Mostafa Akhavan-Safar: Study design, data cleaning and statistical analysis, drafting the manuscript.

Mohammad Mohsen Sadr: interpretation of the results, editing the manuscript.

Mohsen Yoosefi Nejad: Study design, acquisition of data.

All authors read and approved the final manuscript.

### Conflict of interest

The authors declare that no conflicts of interest exist.

## References

- [1] S. S. Virani *et al.*, "Heart disease and stroke statistics-2021 update: a report from the American Heart Association," *Circulation*, vol. 143, no. 8, pp. e254-e743, 2021, <https://doi.org/10.1161/CIR.0000000000000950>.
- [2] K. K. Kapoor, K. Tamilmani, N. P. Rana, P. Patil, Y. K. Dwivedi and S. Nerur, "Advances in social media research: Past, present and future," *Information Systems Frontiers*, vol. 20, pp. 531-558, 2018, <https://doi.org/10.1007/s10796-017-9810-y>.
- [3] M. M. Sadr, A. Mousavi Chelak, S. Ziaei and J. Tanha, "A predictive model based on machine learning methods to recognize fake persian news on twitter," *International Journal of Nonlinear Analysis and Applications*, vol. 11, pp. 119-128, 2020, <https://doi.org/10.22075/IJNAA.2020.4563>.
- [4] H. Anber, A. Salah and A. Abd El-Aziz, "A literature review on Twitter data analysis," *International Journal of Computer and Electrical Engineering*, vol. 8, no. 3, p. 241, 2016, <https://doi.org/10.17706/ijcee.2016.8.3.241-249>.
- [5] P. Hitlin, K. Olmstead, and S. Toor, "Public comments to the federal communications commission about net neutrality contain many inaccuracies and duplicates," *Pew Research Center: Internet, Science & Tech.* United States of America, 2017.
- [6] O. H. A. Mohsen Yoosefi Nejad and M. Hosseinzadeh, *Twitter: The Rich Source of Information*, Amazon, 2022.
- [7] S. A. Brown *et al.*, "Leveraging social media for cardio-oncology," *Current treatment options in oncology*, vol. 21, pp. 1-19, 2020, <https://doi.org/10.1007/s11864-020-00775-3>.
- [8] S. Shadroo, M. Yoosefi Nejad, A. O. Bali, M. Hosseinzadeh and M. S. Delghandi, "A comparison and analysis of the Twitter discourse related to weight loss and fitness," *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 9, pp. 1-12, 2020, <https://doi.org/10.1007/s13721-020-00228-9>.
- [9] M. Yoosefi Nejad, M. S. Delghandi, A. O. Bali and M. Hosseinzadeh, "Using Twitter to raise the profile of childhood cancer awareness month," *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 9, pp. 1-5, 2020, <https://doi.org/10.1007/s13721-019-0206-4>.
- [10] C. N. May *et al.*, "Weight loss support seeking on twitter: the impact of weight on follows back rates and interactions," *Translational behavioral medicine*, vol. 7, no. 1, pp. 84-91, 2017, <https://doi.org/10.1007/s13142-016-0429-1>.
- [11] H. Alhodaib *et al.*, "Evaluation of the Role of Twitter in Increasing the Awareness and Knowledge about Diabetes Among Citizens of Saudi Arabia," *The Open Public Health Journal*, vol. 14, no. 1, pp. 605-611, 2021, <https://doi.org/10.2174/1874944502114010605>.
- [12] M. J. Cumbras-Sánchez, R. Hermoso, D. Iñiguez, J. R. Paño-Pardo, M. Á. A. Bandres and M. P. L. Martinez, "Qualitative and quantitative evaluation of the use of Twitter as a tool of antimicrobial stewardship," *International journal of medical informatics*, vol. 131, p. 103955, 2019, <https://doi.org/10.1016/j.ijmedinf.2019.103955>.
- [13] M. M. Adrian, A. Nurmandi, I. Muallidin and E. P. Purnomo, "Social Network Analysis: How Twitter social media Used in Raising Tax Awareness During the COVID-19 Pandemic," In *International Conference on Human-Computer Interaction*, Cham: Springer International Publishing, 2021, pp. 331-339, [https://doi.org/10.1007/978-3-031-06391-6\\_43](https://doi.org/10.1007/978-3-031-06391-6_43).
- [14] A. Chun, R. Panchmatia, Q. Doan, G. Meckler, B. Narayan and G. D. Meckler, "Twitter as a knowledge translation tool to increase awareness of the OpenHEARTSMAP psychosocial assessment and management tool in the field of pediatric emergency mental health," *Cureus*, vol. 14, no. 8, p. e27597, 2022, <https://doi.org/10.7759/cureus.27597>.
- [15] H. Murray and M. Jackson, M. "Using Twitter Data to Evaluate Pangolin Conservation Awareness," In *The Tenth International Conference on Animal-Computer Interaction*, December 2023, pp. 1-4, <https://doi.org/10.1145/3637882.3637884>.
- [16] C. C. Conley, N. G. Goyal and S. A. Brown, "# CardioOncology: Twitter chat as a mechanism for

- increasing awareness of heart health for cancer patients," *Cardio-oncology*, vol. 6, no. 1, p. 19, 2020, <https://doi.org/10.1186/s40959-020-00072-w>.
- [17] H. Bangash, F. Khan, B. He, M. Arce and I. J. Kullo, "Use of Twitter to promote awareness of familial hypercholesterolemia," *Circulation: Genomic and Precision Medicine*, vol. 12, no. 7, p. e002550, 2019, <https://doi.org/10.1161/CIRCGEN.119.002550>.
- [18] L. Sinnenberg *et al.*, "Twitter as a potential data source for cardiovascular disease research," *JAMA cardiology*, vol. 1, no. 9, pp. 1032-1036, 2016, <https://doi.org/10.1001/jamacardio.2016.3029>.
- [19] D. L. Fischman, M. A. Mamas, M. Alasnag, P. Parwani, M. P. Savage and T. Desai, "Understanding the analytics of Twitter in cardiovascular medicine," *Case Reports*, vol. 2, no. 5, pp. 837-839, 2020.
- [20] M. C. Alraies and S. Sahni, "Why cardiologists should be on social media—the value of online engagement," *Expert Review of Cardiovascular Therapy*, vol. 15, no. 12, pp. 889-890, 2017, <https://doi.org/10.1080/14779072.2017.1408408>.
- [21] M. C. Alraies, S. Raza and J. Ryan, "Twitter as a new core competency for cardiologists," *Circulation*, vol. 138, no. 13, pp. 1287-1289, 2018, <https://doi.org/10.1161/CIRCULATIONAHA.118.032999>.
- [22] J. W. Creswell, "Mixed methods procedures: data collection procedures," *Research Design: qualitative, quantitative, and mixed methods approaches*, pp. 219-220, 2003.
- [23] L. Sinnenberg *et al.*, "Twitter as a potential data source for cardiovascular disease research," *JAMA cardiology*, vol. 1, no. 9, pp. 1032-1036, 2016, <https://doi.org/10.1001/jamacardio.2016.3029>.
- [24] A. K. McCallum, "Mallet: A machine learning for language toolkit (2002)," ed, 2002.



**Mostafa Akhavan-Safar** is an Assistant Professor of Information Technology Engineering currently at the School of Computer and Information Technology Engineering of Payame Noor University (PNU). He received

his M.Sc. in Information Technology Engineering from Iran University of Science and Technology (IUST), and Ph.D. in Information Technology Engineering from Tarbiat Modares University (TMU), Tehran, Iran. His research interests include Bioinformatics, Data Mining, Machine learning, Information systems and Social Network Analysis.



**Mohammad Mohsen Sadr** is an Assistant Professor of Information Technology Engineering currently at the School of Computer and Information Technology Engineering of Payame Noor University (PNU). He received

his M.Sc. in Information Technology Engineering from Tarbiat Modares University (TMU), and Ph.D. in knowledge management from Payame Noor University (PNU), Iran. His research

interests include Business intelligence, artificial intelligence, machine learning and text mining.



**Mohsen Yoosefi Nejad** is an assistant professor in Payame Noor University (PNU) in Iran where he has been a faculty member since 2009. He received his M.Sc. in Information Technology Engineering from Amirkabir University of Technology (AUT), and Ph.D. in computer Engineering from Islamic Azad University (IAU). His research interests include Social Network Analysis, Machine Learning, Image Processing and Quantum Computing.