

Texas A&M University-San Antonio

Digital Commons @ Texas A&M University-San Antonio

Computer Information Systems Faculty
Publications

College of Business

2020

Health Risks of e-cigarettes: Analysis of Twitter Data Using Topic Mining

Abdullah Wahbeh

Slippery Rock University of Pennsylvania

Mohammad A. Al-Ramahi

Texas A&M University-San Antonio, mrahman1@tamusa.edu

Omar El-Gayar

Dakota State University

Tareq Nasralah

Northeastern University

Follow this and additional works at: https://digitalcommons.tamusa.edu/cis_faculty



Part of the [Computer Sciences Commons](#), and the [Data Science Commons](#)

Repository Citation

Wahbeh, Abdullah; Al-Ramahi, Mohammad A.; El-Gayar, Omar; and Nasralah, Tareq, "Health Risks of e-cigarettes: Analysis of Twitter Data Using Topic Mining" (2020). *Computer Information Systems Faculty Publications*. 6.

https://digitalcommons.tamusa.edu/cis_faculty/6

This Conference Proceeding is brought to you for free and open access by the College of Business at Digital Commons @ Texas A&M University-San Antonio. It has been accepted for inclusion in Computer Information Systems Faculty Publications by an authorized administrator of Digital Commons @ Texas A&M University-San Antonio. For more information, please contact deirdre.mcdonald@tamusa.edu.

Health Risks of e-cigarettes: Analysis of Twitter Data Using Topic Mining

Completed Research Full Paper

Abdullah Wahbeh

Slippery Rock University of PA
abdullah.wahbeh@sru.edu

Mohammad Al-Ramahi

Texas A&M University-San Antonio
mohammad.abdel@tamusa.edu

Omar El-Gayar

Dakota State University
omar.el-gayar@dsu.edu

Tareq Nasralah

Northeastern University
t.nasrala@northeastern.edu

Abstract

The recent rise of e-cigarettes and vaping products has increased concerns that another young generation may become addicted to nicotine. Recently, it becomes evident that several health issues are related to the use of e-cigarettes and vaping products. The objective of this paper is to understand and identify such health issues by collecting and analyzing social media data. The analysis reflects the most important themes and topics discussed by online user's about e-cigarettes, vaping, and associated health issues. Using topic modeling techniques, we were able to identify several health issues related to the use of e-cigarettes and vaping products. These issues include lung diseases, coughing and breathing issues, heart related issues, throat burn, respiratory related risks, dizziness, addiction, bronchitis, and cancer.

Keywords

Vaping, e-cigarettes, social media, topic modeling

Introduction

Electronic cigarettes (e-cigarettes), a battery-operated product designed to deliver nicotine, flavor, and other chemicals in aerosol form (Zhan et al. 2017), are widely used around the world. The e-cigarettes market is expected to have an increasing growth rate by 2023, reaching to \$48 billion (Paek et al. 2020).

Nowadays, e-cigarettes are becoming popular product used by millions of adults and youth in the United States and worldwide (Allem et al. 2019). In the United States, the number of youth who used e-cigarettes and never smoked before has increased 3-fold between 2011 and 2013 (Chu et al. 2015). In addition, e-cigarettes use among high school students has increased from 1.5% to 16.0% (Zhou et al. 2018). More than 5 million middle and high school students used e-cigarettes in 2019, and almost one million were using e-cigarette on a daily basis (Cullen et al. 2019; FDA 2020).

Since the introduction of e-cigarettes, awareness concerning the use of such products has risen significantly. The use of e-cigarettes could lead to the initiation of cigarettes use among adolescents, and may recruit youth and non-smokers to nicotine experimentation and dependence (Allem et al. 2019). Despite the fact that some e-cigarette products might be less harmful than combustible products because they do not burn tobacco and produce aerosols with fewer chemicals, e-cigarettes use is not harm free (Hua et al. 2020). The long-term health effects of e-cigarette use are largely unknown (Kim et al. 2017). However, according to the literature, e-cigarette contains nicotine could lead to many health-related issues (Alexander et al. 2019).

Nowadays, social media is considered a promising and viable source of data for gaining insights of various health conditions (Nasralah et al. 2019). However, there is a limited number of studies that utilize social media data to address health related issues associated with the use of e-cigarettes and vaping products, and such studies utilized frequency analysis for identifying such issues. The objective of this research is to investigate e-cigarettes and vaping health related issues by collecting and analyzing recent social media posts. Specifically, we collected tweets that relate to the vaping epidemic from social media users. The collected tweets are analyzed using machine learning techniques to understand the recent themes and perceptions toward the vaping epidemic.

From a theoretical perspective, this study provides a deeper understanding of the overall discussion about vaping and e-cigarettes and highlights important concerns and healthcare related issues of the use of such products comparing to the prevailing beliefs that such products are better than cigarettes and do not cause any healthcare complications. From a practical perspective, given the popularity of social media platforms for information and opinions sharing, it is important to highlight the misconceptions about the use of e-cigarettes and vaping products, which in turns could be used to increase awareness about healthcare related issues associated with vaping and e-cigarettes and defy rumors that such products are not harmful and much better than regular cigarettes.

The remainder of the paper is organized as follows: the next section provides an overview of existing literature related to vaping, e-cigarettes, and social media data analytics. The research design and methodology section discuss data collection, preparation, and analysis. The results and discussion section summarize the findings. The paper concludes with a summary of contributions and limitations.

Literature review

In the literature, a number of studies have addressed several aspects related to vaping and e-cigarettes. At the marketing, commercial, and consumers level, Lazard et al. (2016) addressed the overarching themes and insights for trending topics on e-cigarette from commercial and consumer users. Data were collected using a set of keywords and tweets were then separated into two groups for analysis, tweets and retweets. Authors used text mining to determine patterns and important topics for e-cigarette. Data were processed into tokens and then used to generate a term by frequency matrix for further analysis. Results from topic analysis showed that most of the themes were marketing-focused, where the majority of the remaining themes addressed policies, personal experiences, and the differentiation of e-cigarettes from traditional tobacco. While Paek et al. (2020) investigated marketing and promotional activities about e-cigarettes on social media and how such activities differ between social media platforms. The authors collected and analyzed data from Twitter, news article, and Google trends about e-cigarette. Data was analyzed using frequency analysis, and manual and automatic coding of data into five mutually exclusive themes. Results showed that the most popular themes were product evaluation, promotion, policy, health, and price. The average retweet count of product evaluation tweets was the lowest, where the average retweet count of promotion tweets was the highest. Furthermore, with respect to correlation with the search engine query count, the product evaluation theme count was the highest and the promotion theme count was the lowest. With respect to correlation with online news item count, the health theme count was the highest and the promotion theme count was the lowest. Focusing on the e-cigarette market, Kim et al. (2015) explored marketing trends about e-cigarettes and locations where people use e-cigarettes. The authors collected data from Twitter using a set of keywords, then a random subset was manually labeled as advertising or non-advertising. Classification algorithms were used to code the remaining data. For location information, manual coding and natural language processing were used to indicate locations where e-cigarettes were used. Results showed that the majority of the tweets were advertising, given the fact that most of the tweets were about e-cigarette brands, affiliate marketers, and resellers of e-cigarette products. Finally, only a small number of tweets addressed location information, and findings showed that e-cigarette use was prominent in home/room/bed, school, in public, the bathroom, and at work.

At the flavor, vaping liquid and juice level, E-cigarette flavors have also garnered research attention. For example, Wang et al. (2015) systematically mined e-cigarette and flavor content on Reddit. Data was collected from the top ten popular and relevant subreddits using seven keywords related to electronic cigarette, flavors, and juices. Data were collected based on a set of keywords used in the search and then by ranking the data by relevance, hot spot, importance, up-to-date information, and reply count. Sample of the collected data was used to classify the flavors based on their ingredients. Furthermore, the number of times that each flavor occurred in posts was counted. Finally, characteristics of a variety of flavors were also identified to gain a better understanding regarding flavors popularity. Results showed that fruit flavors were mentioned the most and were popular flavors used in mixed blends and strawberry and vanilla flavors were the most popular for e-juice mixing. Chu et al. (2015) examined message content on Twitter from e-cigarette brands to determine if messages about flavors are more likely than non-flavor messages to be passed along to other viewers. Data were collected from tweets about two e-cigarette brands, Blu and V2, and then processed by identifying messages that contained terms related to e-cigarette flavors. For each tweet, authors counted how many times it was retweeted and collected profile information from the users. A list of search terms was used in automatic content analysis to identify flavor-related messages. A random

sample of the results was checked and validated by a human coder. The authors used chi-square test, t-test, and multiple logistic regression for testing the assumptions. Results showed that flavor-related posts were retweeted at a significantly higher rate by e-cigarette brands and other Twitter users than non-flavor posts. While Zhou et al. (2018) addressed the influence of flavors on e-cigarette related information propagation on social media platform. E-cigarette-related data were collected from Facebook based on a set of keywords generated by a domain expert. A total of eleven binary variables for flavor categories are used to characterize the influence of flavors on information propagation. Topic modeling was used to extract three hidden topics namely, “details about products (product), methods of consuming e-cigarettes (method), and other related discussions (other)”. Regression models and hurdle negative binomial model were used to characterize the influence of different flavors on e-cigarette-related information propagation. Results showed that five flavors, sweet, dessert & bakery, fruits, herbs & spices, and tobacco have negative influences on e-cigarette related information propagation, and that the mention of a flavor in a post does not enhance popularity of e-cigarette related information. Furthermore, some flavors reduced the popularity of information, while a set of non-flavor related factors were associated with information propagation. Also related to flavors, Li et al. (2018) systematically mined users’ e-liquid usage patterns from online reviews using opinion summarization techniques. Online users’ reviews, consisting of the review text, an overall rating and a set of e-liquid aspect ratings were collected from JuiceDB. For data analysis, reviews were labeled positive, neutral, and negative. Data were analyzed using an ID3-based influential aspect analysis model and sentiment analysis. Results showed that flavor accuracy and value were the two most important aspects that affected users’ sentiments toward e-liquids, fruity and sweet were the two most popular flavors, great and sweet tastes, reasonable value, and strong throat hit made users satisfied with fruity and sweet flavors. Furthermore, users complained about some e-liquids’ prices, bad quality, and harsh throat hit.

At the product design level, Allem et al. (2019) studied product design features that may increase the use of e-cigarette. Data related to e-cigarette product reviews was collected from popular e-cigarette-related sub-reddit network. Each review consisted of several sub-sections, related to one product, while others related to all products. Only subsections relating to positive features and negative features in each review were examined. For data analysis and findings, authors read through the most commonly occurring set of words and phrases and identified ten themes. These themes were mainly related to the product’s build quality, color, tip model, battery quality, price, screw quality, power mode performance, coil performance, temperature control performance, and tank quality.

At the e-cigarettes and healthcare level, Hua et al. (2020) extracted and analyzed online forum posts related to e-cigarette and health. A subset of posts related to seven health subforums was selected. Authors annotated the data with medical concepts and analyzed two semantic types, sign or symptoms and disorder or syndrome. Each post was assigned a positive, negative, or neutral sentiment. In order to measure health affects related to electronic cigarettes use, authors used the Random Forest classifier on sample posts that were manually labeled by the authors and predicted the sentiment for unseen posts. Symptom and disorder data were categorized into twelve organ systems or anatomical regions. Results showed that headache, coughing, malaise, asthma, dehydration, and pharyngitis were the most frequent symptoms and disorders.

Several studies have analyzed e-cigarettes related data on the web. These studies addressed aspects of e-cigarettes related to user experience, trials, and adoption (Barrientos-Gutierrez et al. 2019; Chen et al. 2015), products and products design features (Allem et al. 2019; Myslín et al. 2013), flavors and liquids (Chu et al. 2015; Li et al. 2018), online e-cigarettes content and discussion (Glowacki et al. 2017; Kim et al. 2017; Zhan et al. 2017; Zhou et al. 2018), as well as marketing and promotions (Clark et al. 2016; Kim et al. 2015; Lazard et al. 2016; Paek et al. 2020). However, there is a limited number of studies that have addressed health related issues associated with the use of e-cigarettes and vaping products (Hua et al. 2020). Furthermore, such studies have used frequency analysis to determine the most frequently reported symptoms and disorders associated with the use of e-cigarettes.

Research Design and Method

Data Collection and Preparation

The data was collected from Twitter using Crimson Hexagon tool. Crimson Hexagon (CH) is a social media analytics tool employs an unsupervised and supervised machine learning techniques and text analysis model developed by Hopkins and King (2010).

We collected all tweets from 01-01-2019 to 02-15-2020. The time frame was selected because it reflects discussions about recent concerns and news regarding the partial and temporary e-cigarette ban. To retrieve the relevant tweets, we used the search query in Figure 1 that includes several health-related keywords (Hua et al. 2020), such as pain, lung, heart, and cancer. The collected tweets are all selected based on the criteria of having at least one health related keyword and e-cigarette/vaping related keyword, and we excluded retweets. The size of data collected is 69,860 tweets.

To prepare the tweets, we removed numbers, URLs, stop words, and represented each tweet using the well-known Term Frequency Inverse Document Frequency (TF-IDF) weighting scheme (Ramos 2003).

```
(vap* OR E-Cigarette OR e-cigarette OR ecig* OR e-cig* OR (electronic AND cig*))
AND
(pain OR nausea OR dizziness OR dizz* OR lightheadedness OR coughing OR cough* OR wheezing OR wheez*
OR dyspnea OR breathing OR breath* OR heartburn OR Cancer OR cancer OR heart* OR cramping OR cramp*
OR integumentary OR throat OR flatus OR constipation OR voice OR pharyngeal OR itching OR itch* OR
skin OR chestOR fatigue OR tired OR malaise OR disorder* OR symptom OR Symp* OR headache
OR mouth OR neurological OR asthma OR pharyngitis OR dehydration OR dehydr* OR Respiratory OR respir*
OR lung OR doctor OR physician OR digestive OR cold OR flu OR immune OR halitosis OR blood OR COPD
OR chronic OR pneumonia OR bronchitis OR aptyalism OR ulcer OR migraine OR bronchio* OR obliterans)
AND -
(http OR https OR RT OR VICKS OR vick*)
```

Figure 1. Search Query

Topic Modeling

Topic models are types of statistical algorithms for extracting the main topics in a collection of documents. Latent Dirichlet Allocation (LDA) is one of the common topic modeling algorithms used (Blei et al. 2003). The algorithm produces a set of topics with probability distribution over words in each topic. The algorithm also generates probability distributions over topics for each document (Al-Ramahi et al. 2017; Bao and Datta 2014). We used Python to build and run LDA algorithm and identified 50 topics and within each topic showed the top-10 words and their relative weight (i.e. probability). The naming of topics was first conducted by the first author and confirmed by the second author. Naming was initially based on the identification of a logical connection between the 10 most frequent words for a topic.

Results and Discussion

Data analysis using topic modeling revealed a number of health-related issues caused by e-cigarettes and vaping products. These issues were mainly related to the vaping epidemic, lung diseases and injuries, coughing, heart diseases, throat burn, respiratory related risks, dizziness, addiction, stomach issues, bronchitis, and cancer.

Vaping epidemic reflects the popularity of different e-cigarettes and vaping devices among teenagers and youth, and associated healthcare issues. According to Figure 2, tweets related to “epidemic”, “vaping”, “ban”, and “lung” dominated this category. Example tweets include “there’s an epidemic spreading - it’s vaping” and “e-cigarette use among teens is officially an epidemic.”

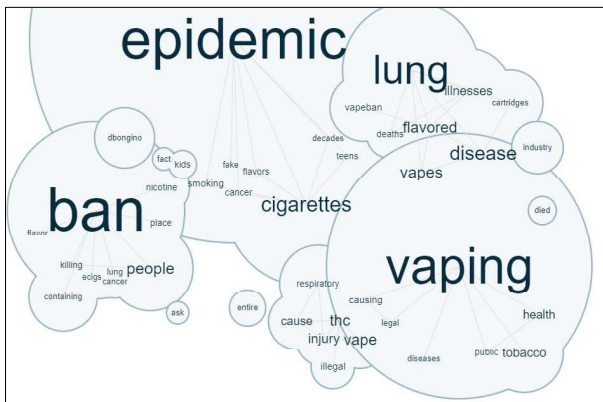


Figure 2. Vaping Epidemic

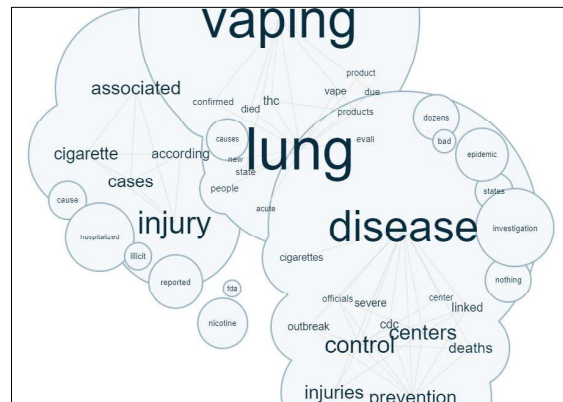


Figure 3. Lung Diseases and Injuries

The use of e-cigarette or vaping products resulted in an outbreak in lung diseases. According to Figure 3, tweets related to “vaping”, “lung”, “disease”, and “injury” dominated this category followed by “centers”, “control”, and ‘deaths”. Example tweets include “*there have been several reports of severe lung disease in young people. Most were using regular vape and marijuana*”, “*this generation seriously created a new type of lung disease from vaping*”, and “*as nicotine is a drug that impacts the cardiovascular systems and the chemicals in vaping solutions are causing lung disease*”. Our findings are in agreements with information available on Centers for Disease Control and Prevention (CDC). In November 2019, CDC has confirmed 2,290 e-cigarette or vaping related lung injury cases, with 47 deaths as of February 4, 2020 (CDC 2020).

Many different reasons could lead to coughing. Vaping techniques and the use of e-cigarettes are considered one of the main reasons for coughing, especially when the body is exposed to vapors from e-cigarettes (Qasim et al. 2017). Furthermore, the use of e-cigarette or vaping is also related to symptoms such as shortness of breath and chest pain (Fumerton et al. 2019). Our findings are in agreements with existing literature. According to Figure 4, tweets related to “cough” and “chest” dominated this category followed by “coughing” and “vaping”. Example tweets include “*if you start coughing get to the doctor fast. My roommate got pneumonia from the vapers causing moisture*” and “*I like vaping, but I started coughing up phlegm with bits of vape oil in it*”.

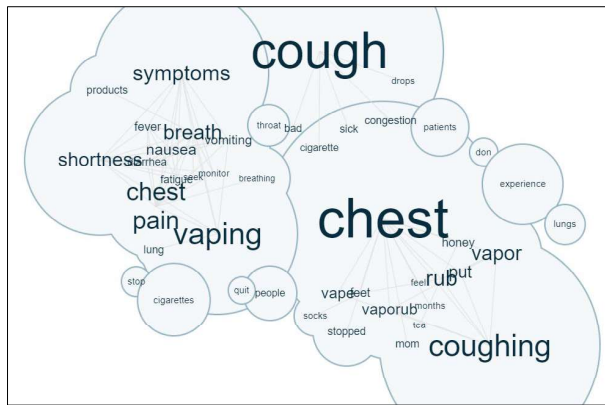


Figure 4. Coughing

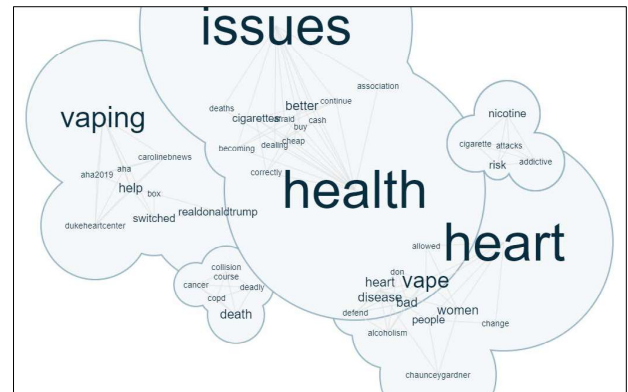


Figure 5. Heart Diseases

The use of e-cigarette or vaping products is related to heart related health issues. The long-term use of e-cigarette or vaping products could cause lethal cardiopulmonary issues such as heart attack, stroke, and congestive heart failure (Perez et al. 2019; Wiebel et al. 2019). Our findings are in agreements with existing literature. According to Figure 5, tweets related to “health”, “heart”, “issues” dominated this category. Example tweets include “*e-cigarette flavors can damage the cells that line your blood vessels and perhaps your heart health down the line*” and “*my chest and lungs and heart were all tight and I had tunnel vision so I am done vaping now*”.

The use of e-cigarette or vaping products is related to throat burn, which could lead to mouth and throat irritation. Such burn could be caused because the nicotine is too strong or if the smoker is new to vaping and has no experience. Our findings are in agreements with existing literature (Balkissoon 2019). According to Figure 6, tweets related to “vapor”, “throat”, and “hot” dominated this category. Example tweets include “*most vape juices make my throat burn*” and “*all the people switch to menthol and start having throat burns from drinking cold liquid after vaping for extended periods of time*”.

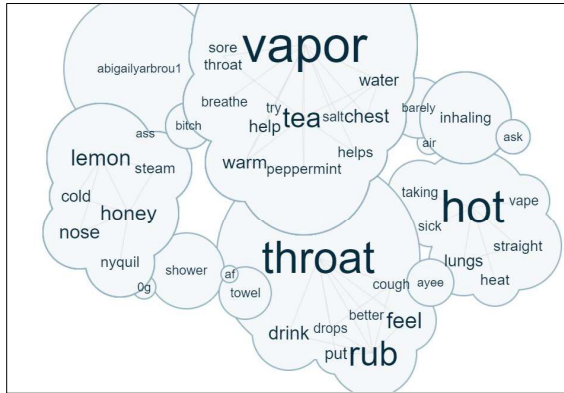


Figure 6. Throat Burn

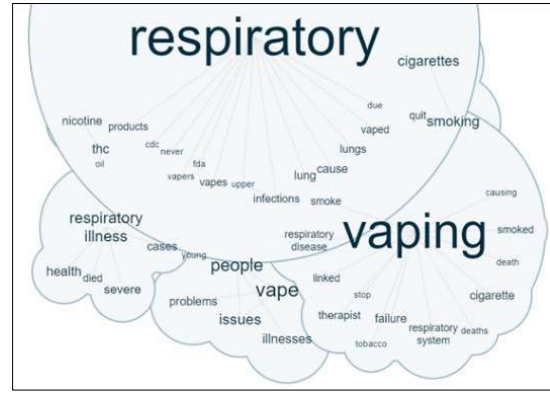


Figure 7. Respiratory Related Risks

The use of e-cigarette or vaping products is related to several respiratory illnesses and diseases. Dicipinigaitis et al. (2019) reported that respiratory failure could happen due to oils inhaled through a variety of vaping products. Respiratory illnesses and diseases caused by vaping products one of the most concerning issues that still need further investigation as quickly as possible (FDA 2020a). Our findings are in agreements with existing literature and information available from the FDA and CDC. According to Figure 7, tweets related to “vaping” and “respiratory” dominated this category. Example tweets include “I can’t be active when my respiratory system is being damaged by unhealthy vapors and or cigarette smoke”, and “black market THC cartridges are what's killing people. It had vitamin E acetate in it, that when inhaled, causes respiratory failure”.

The use of e-cigarette or vaping products is related to dizziness. Dizziness is not mainly related to vaping in general, it is more related to vaping nicotine. According to Hua et al. (2020), dizziness was on the most frequent symptom related to e-cigarette or vaping products usage. Furthermore, according to a number of reports, dizziness was one of the most commonly reported adverse symptoms of e-cigarettes. Our findings are in agreements with existing literature. According to Figure 8, tweets related to “dizzy” followed by “vape” dominated this category. Example tweets include “breathing xylene vapors in small amounts can cause headache, dizziness, drowsiness, and nausea” and “I eventually vaped flavorless nicotine till cigarettes seemed stupid and tasted horrible. A mere 5 years later I got sick of wavering quality and import rules and quit nicotine. Two years of cramps and dizziness. I ain’t going back though”.

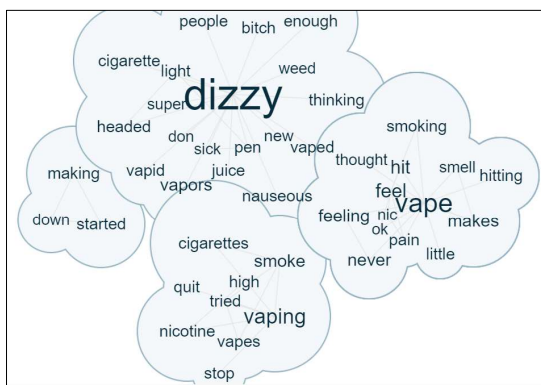


Figure 8. Dizziness

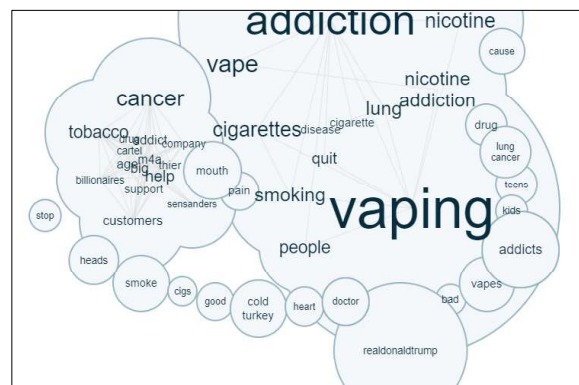


Figure 9. Addiction

The use of e-cigarette or vaping products could lead to addiction-like behaviors by making changes in the symbolic significance and embodied experience of nicotine addiction (Keane et al. 2017). Furthermore, juicy, fruity, or sweet liquid flavor might make users dependent on or be addicted to the product (Li et al. 2018). Our findings are in agreements with existing literature. According to Figure 9, tweets related to “vaping” followed by “addiction” dominated this category. Example tweets include “slurping down 5 coffees a day and vaping to the point of imminent popcorn lung: I'm so glad I've managed to overcome my addictions” and “I think the severity of my respiratory symptoms this time around had convinced me I need to kick the vaping habit, and finally free myself of my nicotine addiction”.

The use of e-cigarette or vaping products could cause upset stomach, a disorder of digestive function characterized by discomfort or heartburn or nausea (Blumenauer et al. 2003). Our findings are in agreements with existing literature. According to Figure 10, tweets related to “stomach” followed by “vaping” dominated this category. Example tweets include “vaping helped a lot & would still be using it, but it hurt my stomach” and “I haven’t had that issue, just the upset stomach when I smoke a higher percent of nicotine”.

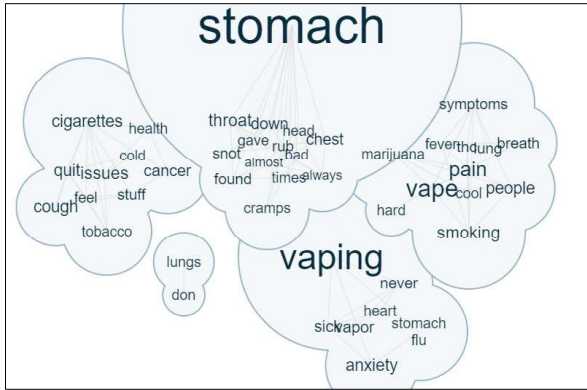


Figure 10. Stomach Issues

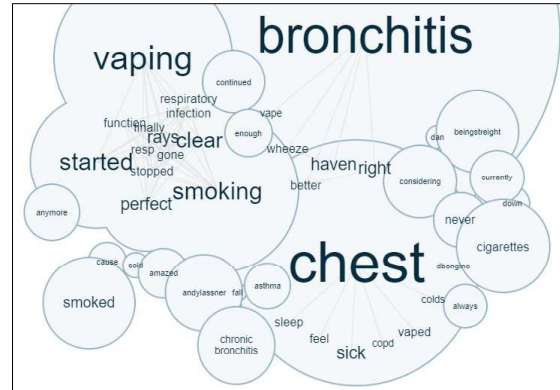


Figure 11. Bronchitis

Chronic bronchitis, sometimes referred to as chest cold, is defined as chronic cough and sputum production for three months a year for two consecutive years (Kim and Criner 2015). Bronchitis causes inflammation of the lining of the bronchial tubes and leads obstruction in airflow (Makh et al. 2018). It could lead to many different consequences including but not limited to hastening lung function decline, exacerbations risks, reducing health related quality of life, and possibly increase mortality. Our findings are in agreements with existing literature. According to Figure 11, tweets related to “bronchitis” and “chest” followed by “vaping” dominated this category. Example tweets include “My doctor was extremely pleased I quit smoking and started vaping. I haven’t been sick or had bronchitis since I’ve been vaping”, “I have bronchitis don’t vape kids”, and “I got chemical bronchitis from an illegal vape”

The use of e-cigarette or vaping products could cause cancer because the vapor from such products contains cancer-causing chemicals. In general, smoking could lead to different kinds of cancers including but not limited to stomach cancer, bladder cancer, throat cancers, and is the cause of nearly 90% of lung cancer cases (Lin 2019). Our findings are in agreements with existing literature. According to Figure 12, tweets related to “vaping” and “lung cancer” followed by “cancer” dominated this category. Example tweets include “suffering from lung cancer after using ‘dank vape cartridges’”, “this would be the second death out of over two hundred cases linked to lung cancer caused by vaping”, and “there have been cases of young people dying in the past month or two in the US from vaping. They are linking it to a case of lung cancer found in a 19yr old boy recently as well”.

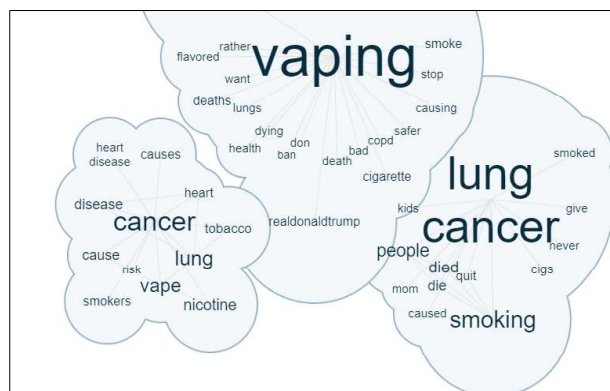


Figure 12. Lung Cancer

Conclusion and Limitations

In this research, we explored health issues related to the use of e-cigarettes and vaping products on social media by collecting and analyzing recent Twitter data. We used machine learning techniques, topic modeling, to automatically analyze the content of the collected tweets. Results showed a number of health-related issues are being discussed on social media. These issues are mainly related to lung diseases and injuries, coughing, heart diseases, throat burn, respiratory related risks, dizziness, addiction, stomach issues, bronchitis, and cancer. Findings from the analysis could help increase the awareness about the vaping epidemic and the health impact of e-cigarettes and vaping products.

This research expanded the existing literature on the use of social media data to address aspects related to e-cigarettes and vaping products. This research addressed an important aspect of e-cigarettes and vaping products, namely the health aspect. While the existing literature focused on aspects related to product design, flavors, marketing, and promotion; limited research addressed health issues related to e-cigarettes and vaping products use. Our findings support existing literature that reported headache, coughing, malaise, asthma, dehydration, and pharyngitis as the most frequently reported symptoms and disorders associated with e-cigarettes. Furthermore, we were able to identify more issues related to dizziness, addiction, stomach issues, bronchitis, and cancer.

The study provides an understanding of information shared on social media platforms, such as Twitter, about vaping and e-cigarettes and highlights important concerns and healthcare related issues associated with such products. Furthermore, it helps fight the misconceptions and defy rumors that such products are not harmful and much better than regular cigarettes. The research could be further improved by identifying a more focused set of tweets by looking at specific age group or products and flavors. Also, additional social media data from other platforms could help improve and strengthen the findings.

References

- Al-Ramahi, M. A., Liu, J., and El-Gayar, O. F. 2017. "Discovering Design Principles for Health Behavioral Change Support Systems: A Text Mining Approach," *ACM Transactions on Management Information Systems (TMIS)* (8:2-3), pp. 1-24.
- Alexander, J. P., Williams, P., and Lee, Y. O. 2019. "Youth Who Use E-Cigarettes Regularly: A Qualitative Study of Behavior, Attitudes, and Familial Norms," *Preventive medicine reports* (13), pp. 93-97.
- Allem, J.-P., Majmundar, A., Dharmapuri, L., Unger, J. B., and Cruz, T. B. 2019. "Insights on Electronic Cigarette Products from Reviews on Therreddit Forum."
- Balkissoon, R. 2019. "Journal Club—Electronic Cigarettes and Vaping as a Harm Reduction Alternative: Really?," *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* (6:3), p. 281.
- Bao, Y., and Datta, A. 2014. "Simultaneously Discovering and Quantifying Risk Types from Textual Risk Disclosures," *Management Science* (60:6), pp. 1371-1391.
- Barrientos-Gutierrez, I., Lozano, P., Arillo-Santillan, E., Morello, P., Mejia, R., and Thrasher, J. F. 2019. "“Technophilia”: A New Risk Factor for Electronic Cigarette Use among Early Adolescents?," *Addictive Behaviors* (91), pp. 193-200.
- Blei, D. M., Ng, A. Y., and Jordan, M. I. 2003. "Latent Dirichlet Allocation," *Journal of machine Learning research* (3:Jan), pp. 993-1022.
- Blumenauer, B. B., Cranney, A., Burls, A., Coyle, D., Hochberg, M. C., Tugwell, P., and Wells, G. A. 2003. "Etanercept for the Treatment of Rheumatoid Arthritis," *Cochrane Database of Systematic Reviews*:3).
- CDC. 2020. "Outbreak of Lung Injury Associated with the Use of E-Cigarette, or Vaping, Products." Retrieved 2/42, 2020, from https://www.cdc.gov/tobacco/basic_information/e-cigarettes/severe-lung-disease.html
- Chen, A. T., Zhu, S.-H., and Conway, M. 2015. "What Online Communities Can Tell Us About Electronic Cigarettes and Hookah Use: A Study Using Text Mining and Visualization Techniques," *Journal of Medical Internet Research* (17:9), p. e220.
- Chu, K.-H., Unger, J. B., Cruz, T. B., and Soto, D. W. 2015. "Electronic Cigarettes on Twitter – Spreading the Appeal of Flavors," *Tobacco Regulatory Science* (1:1), pp. 36-41.
- Clark, E. M., Jones, C. A., Williams, J. R., Kurti, A. N., Norotsky, M. C., Danforth, C. M., and Dodds, P. S. 2016. "Vaporous Marketing: Uncovering Pervasive Electronic Cigarette Advertisements on Twitter," *PLOS ONE* (11:7), p. e0157304.
- Cullen, K. A., Gentzke, A. S., Sawdey, M. D., Chang, J. T., Anic, G. M., Wang, T. W., Creamer, M. R., Jamal, A., Ambrose, B. K., and King, B. A. 2019. "E-Cigarette Use among Youth in the United States, 2019," *Jama* (322:21), pp. 2095-2103.

- FDA. 2020. "Vaporizers, E-Cigarettes, and Other Electronic Nicotine Delivery Systems (Ends)." Retrieved 2/24, 2020, from <https://www.fda.gov/tobacco-products/products-ingredients-components/vaporizers-e-cigarettes-and-other-electronic-nicotine-delivery-systems-ends>
- Fumerton, R., Kling, R., Kim, J., and Gray, A. 2019. "Medical Health Officers' Notice to Report Severe Pulmonary Disease Associated with Vaping or Dabbing," *Cell* (250), pp. 641-1758.
- Glowacki, E. M., Lazard, A. J., and Wilcox, G. B. 2017. "E-Cigarette Topics Shared by Medical Professionals: A Comparison of Tweets from the United States and United Kingdom," *Cyberpsychology, Behavior, and Social Networking* (20:2), pp. 133-137.
- Hopkins, D. J., and King, G. 2010. "A Method of Automated Nonparametric Content Analysis for Social Science," *American Journal of Political Science* (54:1), pp. 229-247.
- Hua, M., Sadah, S., Hristidis, V., and Talbot, P. 2020. "Health Effects Associated with Electronic Cigarette Use: Automated Mining of Online Forums," *Journal of Medical Internet Research* (22:1), p. e15684.
- Kim, A., Hopper, T., Simpson, S., Nonnemaker, J., Lieberman, A., Hansen, H., Guillory, J., and Porter, L. 2015. "Using Twitter Data to Gain Insights into E-Cigarette Marketing and Locations of Use: An Infoveillance Study," *Journal of Medical Internet Research* (17:11), p. e251.
- Kim, A., Miano, T., Chew, R., Eggers, M., and Nonnemaker, J. 2017. "Classification of Twitter Users Who Tweet About E-Cigarettes," *JMIR Public Health and Surveillance* (3:3), p. e63.
- Kim, V., and Criner, G. J. 2015. "The Chronic Bronchitis Phenotype in Copd: Features and Implications," *Current opinion in pulmonary medicine* (21:2), p. 133.
- Lazard, A. J., Saffer, A. J., Wilcox, G. B., Chung, A. D., Mackert, M. S., and Bernhardt, J. M. 2016. "E-Cigarette Social Media Messages: A Text Mining Analysis of Marketing and Consumer Conversations on Twitter," *JMIR Public Health and Surveillance* (2:2), p. e171.
- Li, Q., Wang, C., Liu, R., Wang, L., Zeng, D. D., and Leischow, S. J. 2018. "Understanding Users' Vaping Experiences from Social Media: Initial Study Using Sentiment Opinion Summarization Techniques," *Journal of Medical Internet Research* (20:8), p. e252.
- Lin, W. 2019. "Knowledge and Perceptions of Harm and Interest in Quitting among Cigarette Smokers and E-Cigarette Users: A Cross-Sectional Analysis of the Us Population Assessment of Tobacco and Health Wave 1." The University of Oklahoma Health Sciences Center.
- Makh, P., Mudhalwadkar, R., and Babiwale, A. 2018. "Sensor System Development for Bronchitis Detection from Exhaled Breath," *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT): IEEE*, pp. 1440-1442.
- Myslin, M., Zhu, S.-H., Chapman, W., and Conway, M. 2013. "Using Twitter to Examine Smoking Behavior and Perceptions of Emerging Tobacco Products," *Journal of Medical Internet Research* (15:8), p. e174.
- Nasrallah, T., El-Gayar, O., and Wang, Y. 2019. "What Social Media Can Tell Us About Opioid Addicts: Twitter Data Case Analysis,"
- Paek, H.-J., Baek, H., Lee, S., and Hove, T. 2020. "Electronic Cigarette Themes on Twitter: Dissemination Patterns and Relations with Online News and Search Engine Queries in South Korea," *Health Communication* (35:1), pp. 1-9.
- Perez, M. F., Atuegwu, N. C., Mead, E. L., Oncken, C., and Mortensen, E. M. 2019. "Adult E-Cigarettes Use Associated with a Self-Reported Diagnosis of Copd," *International journal of environmental research and public health* (16:20), p. 3938.
- Qasim, H., Karim, Z. A., Rivera, J. O., Khasawneh, F. T., and Alshbool, F. Z. 2017. "Impact of Electronic Cigarettes on the Cardiovascular System," *Journal of the American Heart Association* (6:9), p. e006353.
- Ramos, J. 2003. "Using Tf-Idf to Determine Word Relevance in Document Queries," *Proceedings of the first instructional conference on machine learning: Piscataway, NJ*, pp. 133-142.
- Wang, L., Zhan, Y., Li, Q., Zeng, D., Leischow, S., and Okamoto, J. 2015. "An Examination of Electronic Cigarette Content on Social Media: Analysis of E-Cigarette Flavor Content on Reddit," *International Journal of Environmental Research and Public Health* (12:11), pp. 14916-14935.
- Wiebel, F., Gohlke, H., and Loddenkemper, R. 2019. "E-Cigarettes: An Underrated Health Risk for Airways, Lung and Cardiovascular System," *German Respiratory Society eV pneumologie. de. doi* (10).
- Zhan, Y., Liu, R., Li, Q., Leischow, S. J., and Zeng, D. D. 2017. "Identifying Topics for E-Cigarette User-Generated Contents: A Case Study from Multiple Social Media Platforms," *Journal of Medical Internet Research* (19:1), p. e24.
- Zhou, J., Zhang, Q., Zeng, D. D., and Tsui, K. L. 2018. "Influence of Flavors on the Propagation of E-Cigarette-Related Information: Social Media Study," *JMIR Public Health and Surveillance* (4:1), p. e27.