Proceedings of the 2023 Design of Medical Devices Conference
DMD2023
April 17-19, 2023, Minneapolis, MN, USA

DMD2023-1316

COUGH AUDIO SENTIMENT ANALYTICS FOR SOFTWARE AS A MEDICAL DEVICE APPLICATIONS

Shivam Damani
Department of Medicine, Mayo Clinic
Rochester, MN

Arshia K. Sethi
Division of Gastroenterology and Hepatology, Department of Medicine, Mayo Clinic
Rochester, MN

Bhavana Baraskar
Department of Radiology, Mayo Clinic
Rochester, MN

Keerthy Gopalakrishnan
Division of Gastroenterology and Hepatology, Department of Medicine, Mayo Clinic
Rochester, MN

Joshika Agarwal
Department of Medicine, Mayo Clinic
Rochester, MN

Hasan A. Albitar
Division of Pulmonary and Critical Care Medicine, Department of Medicine, Mayo Clinic
Rochester, MN

Vaibhav Ahluwalia
Division of Pulmonary and Critical Care Medicine, Department of Medicine, Mayo Clinic
Rochester, MN

Sue Ann P. Donlinger
Division of Pulmonary and Critical Care Medicine, Department of Medicine, Mayo Clinic
Rochester, MN

Vivek Iyer
Division of Pulmonary and Critical Care Medicine, Department of Medicine, Mayo Clinic
Rochester, MN

Shivaram P. Arunachalam
Department of Medicine, Mayo Clinic
Rochester, MN

ABSTRACT

Chronic cough is not only one of the leading causes of seeking healthcare all over the world but also a huge emotional drain on the affected patient population. In this study, we used 24-hour cough recordings to analyze the intervening conversations for sentiment analyses to better diagnose, guide, and manage treatment in such patients. We surveyed a cough clinic and selected four subjects with active cough complaints using relevant ICD-10 codes. Subjects were given and instructed to wear a device to record cough for 24 hours and the recordings were collected at weeks 0, 4, 8, and 12 of the treatment. The collected data was preprocessed to eliminate sections with no data (sleep, silence) and the number of coughs was counted. Google search API calls were used to transcribe the audio files and NLTK's VADER analyzer was used to classify sentiments on a scale of 0 to 1. Finally, average scores were calculated and plotted over a graph to interpret any trends. 12 weeks of cough treatment had varied results on the four subjects. We categorized the exhibited sentiments into negative, neutral, positive, and compound and noted that they also showed no general trends. Among these, the compound sentiment displayed the most erratic patterns, and the obtained results could not generate a steady trend. Further studies are required with a large cohort to collect data over a longer duration to accurately analyze the sentiments associated with chronic cough.

Keywords: Natural Language Toolkit (NLTK), sentiment analysis, Valence Aware Dictionary and Sentiment Reasoner (VADER), chronic cough

1. INTRODUCTION

Cough is an integral part of our essential defense mechanisms, but chronic cough [cough> 8 weeks] is responsible for impairing the quality of life and increasing morbidity in the general population [1]. A study done by Meltzer et al., 2021, showed a 5% prevalence of chronic cough in the United States with a higher prevalence in women [2].

The presence of a chronic persistent cough has a significant impact on the day-to-day life of the patient such as sleep, social activities, and communication. Studies have shown a positive correlation between chronic cough and emotional distress, a 6-16% prevalence of depression and a 9-33% prevalence of anxiety disorders [3], and a fall in depression scores after cough treatment further re-iterate the relationship between symptoms and emotional well-being. The need for cost-effective, non-invasive, patient-centric diagnostic modalities gave rise to the evolution of AI-based analytic methods. In this study, we aim to use the continuous cough recordings and extract intervening conversations, convert them into text and further utilize them in sentiment analyses. Sentiment analysis (SA), commonly referred to as opinion mining, is a subbranch of natural language processing (NLP) that aims to automatically categorize the sentiments conveyed in the free text [4]. Sentiment analyses [SA] emerged with accelerated use of the internet, social networking sites, and blogs and have become invaluable for various businesses, governments, e-commerce websites, and individuals [5]. It is not only applied to product reviews [6], but also to stock markets [7], news articles [8], or political debates [9] to gauge people's opinions [6]. However, recently it gained popularity in healthcare settings to enhance patient care [10].

We propose to use these techniques to guide the diagnoses, management, and prognosis in patients suffering from chronic cough as a readily available software as a medical device (SaMD) product. The lack of any such existing study additionally emphasizes the future utility of this novel approach which can serve as a guide to physicians and in combination with existing treatment methods can significantly improve the quality of life of the patients. Therefore, the purpose of this work was to perform sentiment analysis of audio data from chronic cough patients to better understand the relationship with the treatment approaches that can enhance care management.

2. MATERIALS AND METHODS

2.1 Data collection

The study collected 24-hour recordings of 4 different patients in the cough clinic. The 24-hour recordings were collected at 4 separate times during the subject's treatment – one baseline recording before beginning treatment, one recording 4 weeks from the date of baseline recording, one recording 8 weeks from the date of baseline recording, and one recording 12 weeks from the date of baseline recording. The patients were given a recording device placed appropriately on their bodies to record their coughs for the given duration. The audio data collected was the primary data source for this study. The recording was heard, and a cough count was established.

2.2 Data Pre-Processing

The 24-hour voice recordings were first split at silences, to eliminate all the voice recordings that contained no data (Due to silences, sleep, etc.) These smaller chunks were processed, and the background noise was minimized. Google Search API calls were made to transcribe these audio files, sentence by sentence. NLTK's VADER analyzer was used and the scores of positive, negative neutral, and compound sentiments were found for each sentence. Averages were calculated in each category, and these averages were used to interpret the sentiment of the whole recording. By default, the VADER analyzer ranges compound sentiment between (-1, +1), -1 being negative and +1 being positive. For the results to make sense in the context of this problem, the results were manually shifted from the (-1, +1) to (0, +1) scale.

2.3 Example of sentiment scores

Sentence	*Positive*	*Negative*	*Neutral*	*Compound*
It was on the other side of that wall thing they are more like chairs like this there.	0.0	0.74	0.26	0.679
I dislike this chest sensation a lot, every time I cough it just pains.	0.38	0.63	0.0	-0.660

3. RESULTS AND DISCUSSION

Figure 1 shows the cough count from baseline up to week 12 of the cough treatment for subject 1 superimposed with the sentiment trend. Similarly Figures 2-4 show the cough count from baseline up to week 12 of the cough treatment for subjects 2-4 superimposed with their corresponding sentiment trends. The results show us that the 12 weeks of cough treatment had varied results on the four subjects. Their sentiments also show no general trends.

For subject 1, the graph shows a decrease in cough counts; an increase followed by a decrease in compound sentiment. The positive and negative sentiments followed a similar trend with a rise in week 8 to a decline in week 12. For subject 2, the graph shows an increase followed by a decrease in cough counts; a decrease followed by an increase in compound sentiment. The negative sentiments showed a steady decline till week 12, however, positive sentiments showed an initial decline till week 8 and increased by week 12. For subject 3, the graph shows an increase followed by a decrease in cough counts; a decrease followed by an increase in compound sentiment. The negative sentiments followed a gradually rising pattern throughout the treatment, whereas positive sentiments first raised till week 8 and then dropped by week 12. For subject 4, the graph shows an increase followed by a decrease in cough counts; an increase followed by a decrease in compound sentiment. The positive sentiments steeped till week 8 and then plummeted by week 12. Negative sentiments showed only a slight increase from baseline by week 12 with a minor drop at week 4. These results suggest the possibility of designing cough sentiment analysis tool for a potential software as a medical device (SaMD) product that has wide range utility for patients with chronic cough.

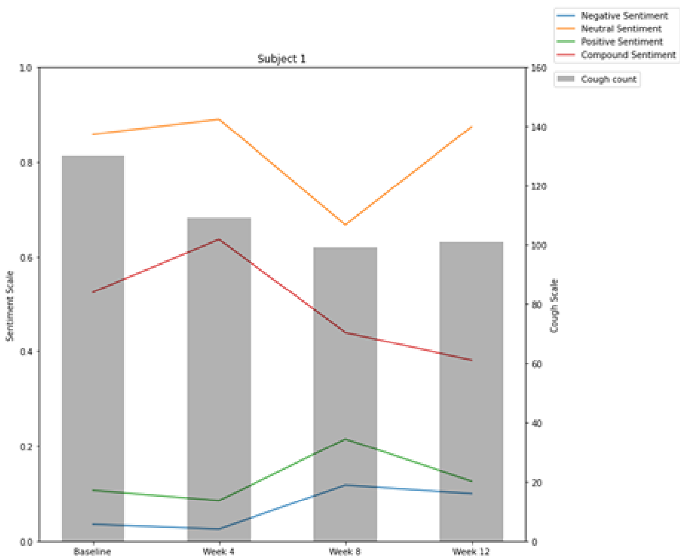

FIGURE 1: Cough counts (bar graph) with the sentiment scores (line graph) for subject 1

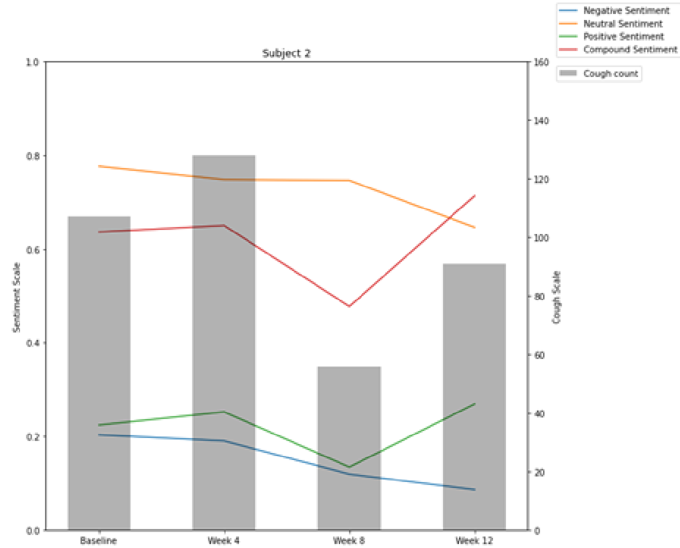

FIGURE 2: Cough counts (bar graph) with the sentiment scores (line graph) for subject 2

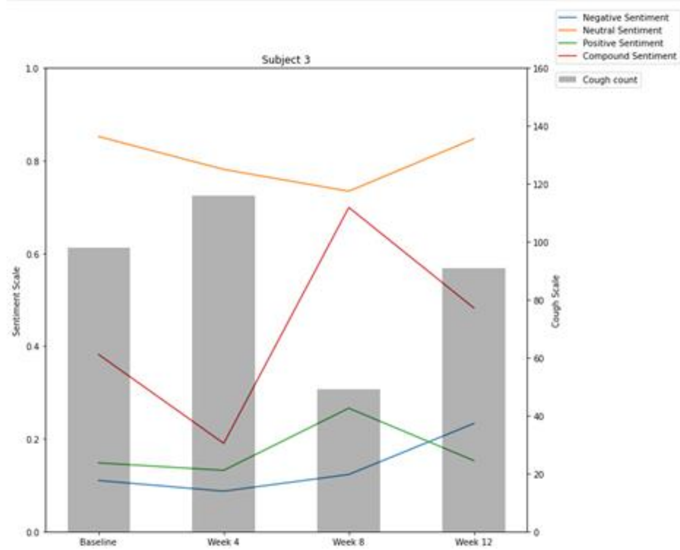

FIGURE 3: Cough counts (bar graph) with the sentiment scores (line graph) for subject 3

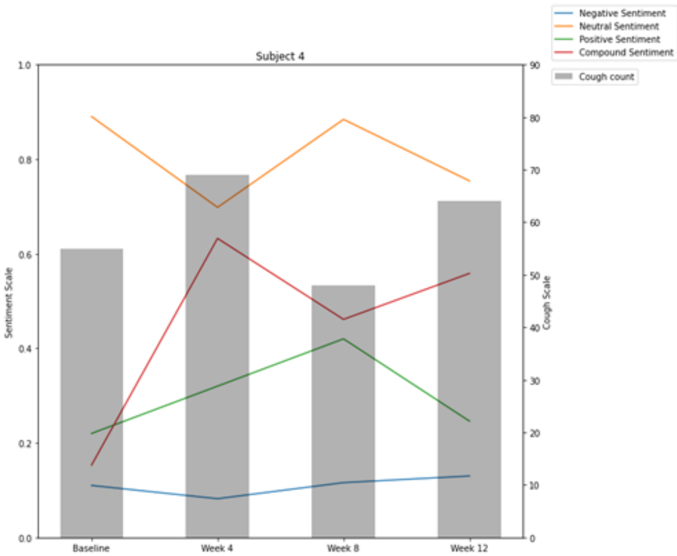

FIGURE 4: Cough counts (bar graph) with the sentiment scores (line graph) for subject 4.

4. CONCLUSION

In this study, the NLTK and the VADER analyzer were applied to conduct a sentiment analysis of speech data of patients with chronic cough and to visualize their sentiments as a function of their cough count. The results indicated that the VADER Sentiment Analyzer was an effective choice for sentiment analysis classification using speech data since it can quickly classify huge amounts of data with potential to design and develop SaMD product for chronic cough clinic applications.

ACKNOWLEDGEMENTS

This work was supported by the Advanced Analytics and Practice Innovation unit for Artificial Intelligence & Informatics research within the Department of Medicine, Mayo Clinic, Rochester, MN USA.

REFERENCES

[1] S. S. Birring, B. Prudon, A. J. Carr, S. J. Singh, M. D. L. Morgan, and I. D. Pavord, “Development of a symptom specific health status measure for patients with chronic cough: Leicester Cough Questionnaire (LCQ),” *Thorax*, vol. 58, no. 4, pp. 339–343, Apr. 2003, doi: 10.1136/thorax.58.4.339.

[2] E. O. Meltzer *et al.*, “Prevalence and Burden of Chronic Cough in the United States,” *The Journal of Allergy and Clinical Immunology: In Practice*, vol. 9, no. 11, pp. 4037-4044.e2, Nov. 2021, doi: 10.1016/j.jaip.2021.07.022.

[3] “Anxiety and depression are associated with reduced quality of life and increased cough severity in chronic cough,” *Asian Pacific Journal of Allergy and Immunology*, 2022, doi: 10.12932/ap-110721-1184.

[4] A. Zunic, P. Corcoran, and I. Spasic, “Sentiment Analysis in Health and Well-Being: Systematic Review,” *JMIR Medical Informatics*, vol. 8, no. 1, p. e16023, Jan. 2020, doi: 10.2196/16023.

[5] D. M. E.-D. M. Hussein, “A survey on sentiment analysis challenges,” *Journal of King Saud University - Engineering Sciences*, vol. 30, no. 4, pp. 330–338, Oct. 2018, doi: 10.1016/j.jksues.2016.04.002.

[6] W. Medhat, A. Hassan, and H. Korashy, “Sentiment analysis algorithms and applications: A survey,” *Ain Shams Engineering Journal*, vol. 5, no. 4, pp. 1093–1113, Dec. 2014, doi: 10.1016/j.asej.2014.04.011.

[7] L.-C. Yu, J.-L. Wu, P.-C. Chang, and H.-S. Chu, “Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock market news,” *Knowledge-Based Systems*, vol. 41, pp. 89–97, Mar. 2013, doi: 10.1016/j.knosys.2013.01.001.

[8] T. Xu, Q. Peng, and Y. Cheng, “Identifying the semantic orientation of terms using S-HAL for sentiment analysis,” *Knowledge-Based Systems*, vol. 35, pp. 279–289, Nov. 2012, doi: 10.1016/j.knosys.2012.04.011.

[9] I. Maks and P. Vossen, “A lexicon model for deep sentiment analysis and opinion mining applications,” *Decision Support Systems*, vol. 53, no. 4, pp. 680–688, Nov. 2012, doi: 10.1016/j.dss.2012.05.025.

[10] A. Zunic, P. Corcoran, and I. Spasic, “Sentiment Analysis in Health and Well-Being: Systematic Review,” *JMIR Medical Informatics*, vol. 8, no. 1, p. e16023, Jan. 2020, doi: 10.2196/16023.