Project 1: Speech Analysis for Medical Diagnosis

Darren Eldredge, *MS ECE Student, Late Entry Accelerated Program (LEAP), College of Engineering, Boston University*

(Submitted to Professor Osama Alshaykh *on September 19, 2021)*

***Abstract*—Medical conditions can alter an individual's voice by affecting organs critical to speech, such as the heart, lungs, brain, muscles, or vocal folds. Vocal signals can be analyzed with Artificial Intelligence and Machine Learning methods to identify vocal biomarkers that can be used for diagnosis, classification, and remote monitoring of patients with a variety of clinical conditions such as Parkinson's Disease, Alzheimer's Disease, Multiple Sclerosis, Type 2 Diabetes, Heart Disease, and COVID-19.**

I. Introduction: Why Your Voice Matters

Public health is a global issue, not wholly owned by any individual, community, or nation. In 2020, Global health spending exceeded $8.3 trillion or 10% of global GDP [1]. Although health service coverage in low- and middle-income countries has shown improvement from 2000 to 2017, it remains well below coverage in wealthier countries [2]. Out-of-pocket spending is also disproportionately higher in low- and middle-income countries [1]. Even a tiny improvement in the accessibility and cost-effectiveness of our healthcare systems and services is therefore a significant opportunity to create wealth and improve equality worldwide.

A study of 17 languages at the University of Lyon in France suggests that no matter what language is used, the human voice conveys information at an average rate of 39.15 bits per second [3], which pales in comparison to the average home internet connection, which has a transfer rate closer to 100 million bits per second. However, the human voice communicates more information than just text, including emotion, identity, and a whole spectrum of biomarkers encoded into the audio signal we produce. These subtle vocal cues may be the key to greater access, affordability, and efficiency of global health assessment technology over the next 10 years, helping to provide earlier and more accurate diagnoses, limit the spread of contagions, decrease costs, and increase the success and survival rates of medical treatment.

We present an overview of applications of voice for health-related purposes, with a focus on COVID-19 testing. This paper includes a literature review, open-source research, and an attempt to duplicate and improve an existing method.

II. Applications and Roadblocks

Rapid advances in natural language processing, audio signal analysis, machine learning, and artificial intelligence are opening the door to the efficient and ethical use of voice applications in healthcare. Work on identifying vocal biomarkers is ongoing for many medical conditions, including Parkinson's Disease, Alzheimer's Disease, Multiple Sclerosis, Rheumatoid Arthritis, Mental Health conditions, Type 2 Diabetes, Coronary Artery Disease, and COVID-19 [4].

While the AI story might seem simple – record speech samples from people with the medical condition as their condition progresses, collect benchmark measurements for progression, use supervised AI techniques to train a model to predict progression from the speech sample – it is not so simple for a few reasons including those listed below [5]:

Training data isn't readily available. Large companies, like Amazon, use hundreds of thousands even millions of hours of unlabeled speech to build acoustic models [6]. There is no data like more data. I have personally reached out to one of the authors of an MIT paper to ask for access to the Open Voice COVID-19 audio samples, and Professor Osama has reached out to Cambridge University to request permission to access their data samples as well.

Data at scale does not exist for some conditions. For some conditions there are not enough patients to generate the amount of data needed to train reliable models, and aggregating data from multiple sources presents its own challenges. Even if the audio data did exist, it is expensive and difficult to generate the clinical benchmarks needed to train the models.

People aren't robots. The characteristics of our speech vary for many reasons. Speech variations may be a function of age, gender, dialect, language, geography, style, etc. and it is important to recognize that even a well-trained AI model may have hidden biases that are unethical and dangerous.

AI models can be difficult to interpret. Mel Frequency Cepstral Coefficients (MFCC) or derivatives may be the key to predicting a disease's progression, but it is difficult to define and understand why that data is clinically relevant.

Incorrect predictions in healthcare applications have a direct impact on the health and safety of patients.

III. Area of Focus: COVID-19

Since the start of the global COVID-19 pandemic, there have been 27,451,761 molecular tests administered in Massachusetts to diagnose infection with SARS-CoV-2 [7]. Assuming a cost of 10.8 USD per RT-PCR test [8], COVID-19 screening has laden taxpayers, institutions, and insurance providers in this state with a bill of nearly $300,000,000. In the week between Sep 10 and Sep 17, there were over 17 million tests administered in the US alone [9]. While we hope for herd immunity through vaccination in developed countries like the United States, thereby diminishing the need for testing, low-income countries are far behind.

During the pandemic, it has become a focus of researchers to try to distinguish and differentiate symptoms of COVID-19 from similar diseases such as influenza. The general diagnostic method is the PCR test, or at Boston University, the RT-PCR test. The mandatory visit to a hospital, clinic, or other testing site is counterproductive, and endangers patients and healthcare professionals alike.

Given the challenges of applying speech analysis in healthcare applications, COVID-19 diagnosis seems like the most promising and least difficult place to start. It is a highly visible, urgent, and important issue to every nation and community on the planet. Importantly, testing is ubiquitous, and a painful, expensive, time-consuming process for many individuals and organizations.

IV. Literature Review

For this review, I searched Google for reliable sources of information on speech analysis, speech analysis for medical diagnosis, and audio signal analysis for COVID diagnosis. My starting point was an article titled Voice for Health: The Use of Vocal Biomarkers from Research to Clinical Practice [4]. This article provides an excellent overview of the topic, from the definition of a vocal biomarker to applications in modern medicine. The article contains a list of conditions suitable for research, and it also explains some very high-level machine learning tactics currently deployed in research. From there, I sought data specific to the global COVID-19 pandemic from the U.N., the W.H.O., the CDC, Massachusetts, Boston University, and several articles from PubMed.org. These articles contained information about the severity and spread of COVID variants, as well as the cost and effectiveness of various types of COVID-19 testing and treatment.

I then began searching for data and open-source projects specifically tied to audio signal analysis and COVID-19 testing. I found COVID-19 breathing and coughing sounds collected for academic research purposes only by the University of Cambridge [10], and I reached out to my professor to try to obtain that data. I reached out to MIT to ask about their Open Access research and to try to obtain the data from the Open Voice COVID-19 initiative [11]. I found the article “Identifying COVID-19 by using spectral analysis of cough recordings: a distinctive classification study” to be very insightful [12]. Features were extracted from using Mel frequency cepstral coefficients (MFCC) and power spectral density based on short-time Fourier transform. The researchers used a support vector machine algorithm (SVM) to the processed signals to identify and classify COVID-19 cough. The cough of subjects with COVID-19 was detected with 95.86% classification accuracy with the proposed method [12]. The diagnosis of COVID-19 coughs was performed with 98.6% and 91.7% sensitivity and specificity, respectively [12]. Sensitivity here meaning a test’s ability to designate an individual with COVID-19 as positive. Specificity here meaning a test’s ability to designate an individual who does not have a disease as negative. For reference, the RT-PCR test for COVID-19 has a sensitivity >99% [13]. However clinical performance of testing depends on a variety of factors and can fall as low as 80% sensitivity and 98-99% specificity [13]. These, then are the targets to beat for any AI enabled method of COVID-19 testing. A key question I took away from this article is whether AI enabled methods can attain similarly high sensitivity and specificity for asymptomatic patients.

The next study I found focused on different types of cough and distinguished patients with COVID-19 from patients with asthma, bronchitis, and healthy individuals with 95.04% accuracy [14]. These articles led me to start looking for open-source data and projects related to COVID-19 audio signal analysis.

Here are the best articles I found for deep understanding of COVID sound analysis with ML and AI models.

VIRUFY_HARVARD: https://ui.adsabs.harvard.edu/abs/2021arXiv210301806F/abstract

Summary: multi-branch deep learning network trained and tested on crowdsourced data, not manually processed, or cleaned. (AUC of .99 for audio samples with COVID-19 positive labels) [15].

MIT_IEEE: https://ieeexplore.ieee.org/document/9208795

Summary: CNN-based model trained on 4256 subjects which diagnoses COVID-19 with a sensitivity of 98.5%, specificity 94.2%. For asymptomatic subjects it achieves sensitivity of 100% with specificity of 83.2% [16].

WOLFRAM_RESEARCH: https://community.wolfram.com/groups/-/m/t/2166833

Summary: Recurrent Neural Network, MFCC feature extraction. Accuracy around 86% with extremely limited data (121 samples) [17].

V. Open Source Research

There are several open-source data sets being established, including MIT’s Open Voice COVID-19 Cough dataset.

My open-source research included the two links provided in class [18][19]:

https://github.com/varmichelle/Diagnosing-Parkinsons

https://www.kaggle.com/paultimothymooney/medical-symptoms-text-and-audio-classification

There are several open-source projects for audio signal analysis, like Deep Speech [21]. However, it seems that the race is on to develop new testing methods for COVID-19 and most organizations, while willing to share data, are less willing to share code. I hope that continued searching will yield more results.

References

[1] Who.int. 2021. Global spending on health: Weathering the storm. [online] Available at: <https://www.who.int/publications/i/item/9789240017788> [Accessed 19 September 2021].

[2] “Health,” United Nations. [Online]. Available: https://www.un.org/en/global-issues/health. [Accessed: 19-Sep-2021].

[3] Science.org. 2021. Human speech may have a universal transmission rate: 39 bits per second. [online] Available at: <https://www.science.org/news/2019/09/human-speech-may-have-universal-transmission-rate-39-bits-second> [Accessed 19 September 2021].

[4] Fagherazzi G, Fischer A, Ismael M, Despotovic V: Voice for Health: The Use of Vocal Biomarkers from Research to Clinical Practice. Digit Biomark 2021;5:78-88. doi: 10.1159/000515346

[5] Aural Analytics. 2021. Vocal Biomarkers: Tempering the AI Hype | Aural Analytics. [online] Available at: <https://auralanalytics.com/vocal-biomarkers-tempering-the-ai-hype/> [Accessed 19 September 2021].

[6] Parthasarathi, Sree Hari Krishnan, and Nikko Strom. "Lessons from building acoustic models with a million hours of speech." ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019.

[7] Mass.gov. 2021. COVID-19 Response Reporting. [online] Available at: <https://www.mass.gov/info-details/covid-19-response-reporting> [Accessed 18 September 2021].

[8] Jakobsen KK, Jensen JS, Todsen T, Tolsgaard MG, Kirkby N, Lippert F, Vangsted AM, Martel CJ, Klokker M, von Buchwald C. Accuracy and cost description of rapid antigen test compared with reverse transcriptase-polymerase chain reaction for SARS-CoV-2 detection. Dan Med J. 2021 Jun 14;68(7):A03210217. PMID: 34169830.

[9] Centers for Disease Control and Prevention. 2021. COVID Data Tracker. [online] Available at: <https://covid.cdc.gov/covid-data-tracker/#cases_tests7day> [Accessed 18 September 2021].

[10] Covid-19-sounds.org. 2021. Covid-19 Sounds App - University of Cambridge. [online] Available at: <https://www.covid-19-sounds.org/en/> [Accessed 18 September 2021].

[11] Opensigma.mit.edu. 2021. MIT Covid-19 Initiative. [online] Available at: <https://opensigma.mit.edu/> [Accessed 20 September 2021].

[12] Melek Manshouri, N. Identifying COVID-19 by using spectral analysis of cough recordings: a distinctive classification study. Cogn Neurodyn (2021). https://doi.org/10.1007/s11571-021-09695-w

[13] College of American Pathologists. 2021. How Good are COVID-19 (SARS-CoV-2) Diagnostic PCR Tests?. [online] Available at: <https://www.cap.org/member-resources/articles/how-good-are-covid-19-sars-cov-2-diagnostic-pcr-tests> [Accessed 18 September 2021].

[14] Pal A, Sankarasubbu M (2021) Pay attention to the cough. In: Proc. 36th Annu. ACM Symp. Appl. Comput., ACM, New York, NY, USA, pp. 620–628. 10.1145/3412841.3441943

[15] Fakhry, A., Jiang, X., Xiao, J., Chaudhari, G., Han, A., and Khanzada, A., “Virufy: A Multi-Branch Deep Learning Network for Automated Detection of COVID-19”, <i>arXiv e-prints</i>, 2021.

[16] Laguarta, Jordi et al. "COVID-19 Artificial Intelligence Diagnosis Using Only Cough Recordings." IEEE Open Journal of Engineering in Medicine and Biology 1 (September 2020): 275 - 281.

[17] Community.wolfram.com. 2021. Predicting COVID-19 using cough sounds classification - Online Technical Discussion Groups—Wolfram Community. [online] Available at: <https://community.wolfram.com/groups/-/m/t/2166833> [Accessed 18 September 2021].

[18] GitHub. 2021. GitHub - varmichelle/Diagnosing-Parkinsons: Synopsys and CSSF project: Classification and diagnosis of Parkinson's Disease. [online] Available at: <https://github.com/varmichelle/Diagnosing-Parkinsons> [Accessed 18 September 2021].

[19] Kaggle.com. 2021. Medical Symptoms Text and Audio Classification. [online] Available at: <https://www.kaggle.com/paultimothymooney/medical-symptoms-text-and-audio-classification> [Accessed 18 September 2021].

[20] Chloë Brown, Jagmohan Chauhan, Andreas Grammenos, Jing Han, Apinan Hasthanasombat, Dimitris Spathis, Tong Xia, Pietro Cicuta, and Cecilia Mascolo. 2020. Exploring Automatic Diagnosis of COVID-19 from Crowdsourced Respiratory Sound Data. Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Association for Computing Machinery, New York, NY, USA, 3474–3484. DOI:https://doi.org/10.1145/3394486.3412865

[21] GitHub. 2021. GitHub - mozilla/DeepSpeech: DeepSpeech is an open source embedded (offline, on-device) speech-to-text engine which can run in real time on devices ranging from a Raspberry Pi 4 to high power GPU servers.. [online] Available at: <https://github.com/mozilla/DeepSpeech> [Accessed 18 September 2021].

[22] Schellekens, I., 2021. COVID-19 is a developing country pandemic. [online] Brookings. Available at: <https://www.brookings.edu/blog/future-development/2021/05/27/covid-19-is-a-developing-country-pandemic/> [Accessed 18 September 2021].

[23] https://drive.google.com/drive/u/0/folders/1kONSI53BvAAd7TvgNGzNAP8dT-Oq3iW1

[24] https://www.worldbank.org/en/news/podcast/2021/07/30/-absolutely-unacceptable-vaccination-rates-in-developing-countries-the-development-podcast