Recent Advances in Computer Audition for Diagnosing COVID-19: An Overview

Kun Qian*, Björn W. Schuller†, and Yoshiharu Yamamoto*

*Educational Physiology Laboratory, The University of Tokyo, Japan, Email: {qian, yamamoto}@p.u-tokyo.ac.jp

†GLAM – Group on Language, Audio, & Music, Imperial College London, UK,

Email: bjoern.schuller@imperial.ac.uk

Abstract—**Computer audition (CA) has been demonstrated to be efficient in healthcare domains for speech-affecting disorders (e.g., autism spectrum, depression, or Parkinson's disease) and body sound-affecting abnormalities (e.g., abnormal bowel sounds, heart murmurs, or snore sounds). Nevertheless, CA has been underestimated in the considered data-driven technologies for fighting the COVID-19 pandemic caused by the SARS-CoV-2 coronavirus. In this light, summarise the most recent advances in CA for COVID-19 speech and/or sound analysis. While the milestones achieved are encouraging, there are yet not any solid conclusions that can be made. This comes mostly, as data is still sparse, often not sufficiently validated and lacking in systematic comparison with related diseases that affect the respiratory system. In particular, CA-based methods cannot be a standalone screening tool for SARS-CoV-2. We hope this brief overview can provide a good guidance and attract more attention from a broader artificial intelligence community.**

Index Terms—**Computer audition, COVID-19, diagnosis, machine learning, overview.**

I. INTRODUCTION

At the time of writing, the COVID-19 pandemic (caused by the SARS-CoV-2 coronavirus) has affected more than 200 countries and regions with more than 63 million confirmed cases, leading to more than 1.4 million deaths (reported by Johns Hopkins Coronavirus Resource Centre, Baltimore, MD, USA). To fight this global crisis, we have witnessed a plethora of endeavours made by a broad scientific community including biology, medicine, sociology, and the information communication technology (ICT). In particular, advanced technologies in the 5 G, internet of things (IoT), and artificial intelligence (AI) have been successfully applied to the diagnosis, management, monitoring, and tracking of COVID-19. Among these AI based applications, fast screening of the chest CT has shown promising results in the current state-of-the-art works. Nevertheless, computer audition (CA) has been underestimated even though it has been increasingly studied in the domain of healthcare [1].

To the best of our knowledge, the main authors of this contribution firstly presented the idea of leveraging the power of CA to combat this war between the virus and humans [2]. In that opinion paper, we systematically summarised the milestones achieved by advanced CA methods and their applications in speech and sound analysis, which can be possibly used in the COVID-19 context. Furthermore, we showed a preliminary experimental study on analysing the spontaneous speech sound of COVID-19 patients collected from Wuhan city, China [3]. In the past eight months, we have seen an increasing number of literature focusing on CA methods for fighting the COVID-19 pandemic [4]–[8]. However, an overview of these existing studies is lacking, which restrains the sustainable development and guidance to the ongoing relevant works. To this end, we write this brief overview article to summarise the current achievements and indicate future directions of CA for COVID-19 studies.

The main contributions of this work are: First, we give a concise summary of the state-of-the-art works on using CA methods for combating the COVID-19 pandemic. We investigate and compare the models by indicating their main methods and results in each literature contribution. Second, we provide an assessment to the readers to what extent CA can be exploited in practice in this global crisis given the encouraging results already achieved. On the other hand, we point out the limitations of the current studies, which should be borne on mind, and reasonably solved in future works. Third, we show future research directions. The remainder of this paper will be laid out as follows: We firstly make a brief description of the existing databases and methods in Section II. Then, we present the current milestones reported in the ongoing studies in Section III. Finally, we conclude this overview article in Section IV.

II. DATABASES AND METHODS

Table I proposes a list of different methods and models used in the literature. Considering the database (including breath, cough, and speech from both healthy control and COVID-19 patients), most of the studies are still ongoing collection work. Most authors claimed in their work that the databases will be released for public research usage in the near future.

III. CURRENT FINDINGS

Generally, the current results are encouraging. However, we cannot make direct comparisons due to the different tasks, data sets, and annotation approaches. Moreover, some of the studies are lacking rigid ground-truth annotation (most of them were based on self-reported symptoms or confirmation rather than chest CT check or RT-PCR test). The *data scarcity* is still the first challenge in the current studies. This holds in particular for highly validated data, but also for highly diverse control-group data to ensure it includes challenging other cases – for example, other diseases affecting the respiratory

TABLE I

Published Literature on Computer Audition for COVID-19. MFCCs: Mel-frequency Cepstral Coefficients. GFW: Glottal Flow Waveform. SVM: Support Vector Machine. PCA: Principal Component Analysis. LR: Logistic Regression. CNN: Convolutional Neural Network. GRU-RNN: Gated Recurrent Unit based Recurrent Neural Network. DT: Decision Tree. RF: Random Forest. AB: AdaBoost. CL: Chance Level. CV: Cross Validation. LOSO-CV: Leave-One- Subject-Out Cross Validation. Sev.: Severity. S: Sleep Quality. F: Fatigue. A: Anxiety.

Ref.	Methods	# Subjects	# Instances	Evaluation	Results
[3]	Hand-crafted features, e. g., MFCCs, F0, SVM as the classifier	51	260	LOSO-CV	**CL**: 33.3 %, UAR: 68.0 % for **Sev** task, 61.0 % for **S** task, 46.0 % for **F** task, 56.0 % for **A** task
[4]	Hand-crafted features & VGGish learnt features, LR, GBT, and SVM as the classifiers	282, 52, 41	439, 86 74	User-based CV	AUC: 0.80, 0.82, 0.80
[5]	MFCCs, Spectrogram, PCA, CNN, SVM	n/a	543	5-fold CV	Accuracy: 88.8 %
[6]	Spectrogram CNN (ResNet-18)	1 039	3 117	Train/ Validation	AUC: 0.72
[7]	MFCCs CNN (ResNet-50)	5 320	n/a	Train/ Validation	AUC: 0.97
[8]	Attention-based Transformer, GRU-RNNs, SVM	88	292	LOSO-CV, 5-fold CV	F1: 0.74 to 0.80, w transformer pre-training F1: 0.67 to 0.70, w/o transformer pre-training
[9]	Vocal Fold Oscillation Coefficients, LR, SVM, DT, RF, AB	19	3 835	3-Fold CV	AUC: 0.83
[10]	GFW, Attention-based CNN	19	3 835	3-Fold CV	AUC: 0.85
[11]	Mel-Filter Bank Features , Phoneme Posteriors, SVM	19	702	6-fold CV	Accuracy: 88.6 %, F1: 0.93

system. Another general discussion is whether asymptomatic COVID-19 affected individuals can be diagnosed by audio approaches. The authors in [8] believe so, but this may also be a discussion of the definition of 'asymptomatic'. If there are no symptoms at all, it seems unclear how the vocal production mechanisms or coughing should be affected. In addition, there is increasing evidence that COVID-19 can go without any affection of the respiratory systems, such as purely inducing diarrhea. It is unclear to which degree such cases will impact. Most importantly, we will need to gain deeper insight into features that are characteristic and the reasoning behind. Current reports are limited, for example, to description of the vocal fold oscillation patterns, or spectral changes. However, higher level impacts such as behavioural change markers should also be taken more into account. In particular, most studies so far compare different subjects for cases of COVID-19 or non-COVID-19. However, more insights will be needed comparing data from the same subjects in both states. Sensing the human mental and/or physical health status can usually not be perfectly acquired by only one modality, e. g., audio. Therefore, we believe by introducing more data modalities such as those accessible by wearable devices, can contribute to a higher performance of the models.

IV. Conclusion

In this brief overview, we proposed the achievements and limitations in the ongoing studies on CA for fighting the COVID-19. The existing literature showed encouraging results, whereas further deeply studies are still needed. As a non-invasive approach, we believe CA based models can contribute to this battle between human and virus given its many advantages such as contact-less sensing and easy to spread at low cost possible distribution as well as real-time assessment.

References

[1] K. Qian *et al.*, "Computer audition for healthcare: Opportunities and challenges," *Frontiers in Digital Health*, vol. 2, p. 5, 2020.

[2] B. W. Schuller *et al.*, "COVID-19 and computer audition: An overview on what speech & sound analysis could contribute in the SARS-CoV-2 corona crisis," *arXiv preprint arXiv:2003.11117*, pp. 1–7, 2020.

[3] J. Han *et al.*, "An early study on intelligent analysis of speech under COVID-19: Severity, sleep quality, fatigue, and anxiety," in *Proc. INTERSPEECH*, Shanghai, China, 2020, pp. 4946–4950.

[4] C. Brown *et al.*, "Exploring automatic diagnosis of COVID-19 from crowdsourced respiratory sound data," *arXiv preprint arXiv:2006.05919*, pp. 1–9, 2020.

[5] A. Imran *et al.*, "AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app," *Informatics in Medicine Unlocked*, pp. 1–27, 2020, accepted, in press.

[6] P. Bagad *et al.*, "Cough against COVID: Evidence of COVID-19 signature in cough sounds," *arXiv preprint arXiv:2009.08790*, pp. 1–12, 2020.

[7] J. Laguarta, F. Hueto, and B. Subirana, "COVID-19 artificial intelligence diagnosis using only cough recordings," *IEEE Open Journal of Engineering in Medicine and Biology*, pp. 2047–2051, 2020.

[8] G. Pinkas *et al.*, "SARS-CoV-2 detection from voice," *IEEE Open Journal of Engineering in Medicine and Biology*, vol. 1, pp. 268–274, 2020.

[9] M. A. Ismail, S. Deshmukh, and R. Singh, "Detection of COVID-19 through the analysis of vocal fold oscillations," *arXiv preprint arXiv:2010.10707*, pp. 1–5, 2020.

[10] S. Deshmukh, M. A. Ismail, and R. Singh, "Interpreting glottal flow dynamics for detecting COVID-19 from voice," *arXiv preprint arXiv:2010.16318*, pp. 1–5, 2020.

[11] K. V. S. Ritwik, S. B. Kalluri, and D. Vijayasenan, "COVID-19 patient detection from telephone quality speech data," *arXiv preprint arXiv:2011.04299*, pp. 1–6, 2020.