2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Novel COVID-19 Screening Using Cough Recordings of A Mobile Patient Monitoring System

Xiyu Zhang*, Michael Pettinati, Ali Jalali, Kuldeep Singh Rajput, and Nandakumar Selvaraj

Abstract— **Since the COVID-19 pandemic began, research has shown promises in building COVID-19 screening tools using cough recordings as a convenient and inexpensive alternative to current testing techniques. In this paper, we present a novel and fully automated algorithm framework for cough extraction and COVID-19 detection using a combination of signal processing and machine learning techniques. It involves extracting cough episodes from audios of a diverse real-world noisy conditions and then screening for the COVID-19 infection based on the cough characteristics. The proposed algorithm was developed and evaluated using self-recorded cough audios collected from COVID-19 patients monitored by Biovitals® Sentinel remote patient management platform and publicly available datasets of various sound recordings. The proposed algorithm achieves a duration Area Under Receiver Operating Characteristic curve (AUROC) of 98.6% in the cough extraction task and a mean cross-validation AUROC of 98.1% in the COVID-19 classification task. These results demonstrate high accuracy and robustness of the proposed algorithm as a fast and easily accessible COVID-19 screening tool and its potential to be used for other cough analysis applications.**

Index Terms— **Machine learning, Signal processing, Audio Analysis, COVID-19 screening, Convolutional neural network (CNN).**

I. INTRODUCTION

The health care systems across the world constantly endeavor to allow effective testing, inoculating and treating for COVID-19 pandemic in an unprecedented massive scale, but many parts of the world are still experiencing or being threatened with more waves or surges of new COVID-19 cases [1]. Reverse transcriptase polymerase chain reaction (RT-PCR), being by far the most accurate testing methods for COVID-19, is limited by its narrow testing capacity. These standard tests are laborious, time consuming, costly and also not easily accessible, especially for the developing countries across the world. Furthermore, the false negative rate of the RT-PCR is relatively high during the course of infection and reach to a lowest of 20% on day 8 of infection [2]. The downsides of the current testing methods and the lack of clinical evidence call for prompt efforts to develop better alternative technologies for timely, accurate, accessible and widespread screening of COVID-19 suspected patients in order to timely treat the COVID-19 patients with appropriate antiviral, monoclonal antibody or other emerging therapies and contain the spread of COVID-19 globally.

The latest literature showcases the use of Artificial-Intelligence (AI) techniques to identify COVID-19 with high level of accuracy using simple cough sound recordings, which can be a fast and inexpensive alternative to the current testing methods [3]–[6]. However, the audio samples used in these studies are usually recorded in a controlled environment such that they are only comprised of cough events with minimal ambient sounds or background noises. In a real-world use-case scenario, the audio recordings are often mixed with a variety of noise and uncontrolled sound signals such as sneezes, clearing throat, speech, television sounds, laughing, tapping, etc., that patients themselves often make or encounter from continuous interactions with the surroundings and electronic devices from their day-to-day lives. Therefore, it is important to build machine learning models to be robust enough for noisy practical conditions.

Besides the noise handling, audio segmentation and accurate extraction of cough episodes is a pivotal algorithmic step that can characterize the cough signatures and effectively quantify associated metrics such as cough frequency, cough duration, cough intensity, etc. There have been some research reports on cough segmentation and extraction [7], [8]. However, their algorithmic performances have been often reported based on simple counting of cough episodes and evaluation metrics for classification tasks rather than conducting systematic performance evaluations involving episodic and duration performance metrics.

In overcoming the present limitations, the study proposes a novel algorithmic framework involving a set of unique modules of noise reduction, cough extraction localizing cough onsets and offsets, and a low-complexity convolutional neural network (CNN) model for both the detection of cough episodes and diagnosis of COVID-19 from the detected coughs. The algorithm performances have been evaluated using audio recordings of the Biovitals® Sentinel smartphone application that included real-world coughs from confirmed COVID-19 patients and also publicly available diverse datasets of various sound recordings.

Furthermore, the algorithm discussed in this paper serves as one component of a larger system designed to analyze cough audios from COVID-19 patients. This overall system is configured with a three-phase cascading architecture:

1) Pre-screen each audio file to identify whether it contains any cough sound.
2) If cough is determined to be present in the audio file, remove noise, extract the cough episode(s) from the audio file.
3) Input the extracted coughs into a cough analysis model to determine if the patient is COVID-19 positive.

Phase 1 in this pipeline is reported in a concurrent submission to the EMBS conference [9]. In this paper, we

Biofourmis Inc, Boston, MA 02110, USA.
e-mail: sylvia.zhang@biofourmis.com

mainly focus on discussing the phase 2 and 3 models. In the following sections we present the data collection (Section II-A), the audio noise reduction (Section II-B), the manual labelling (Section II-C) and the cough extraction process (Section II-D and II-E). The COVID-19 classification model in phase 3 is presented in Section II-F. The performance analysis results of the above models and algorithms are then reported and discussed in Section III and IV and concluded in Section V.

II. METHODS AND MATERIALS

A. Data Collection

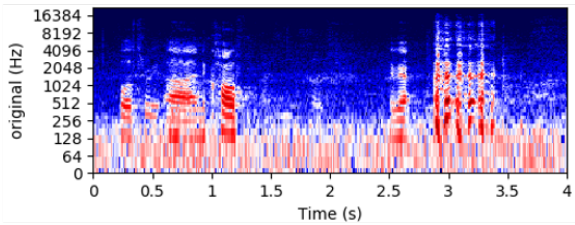
Our proposed system is built on top of Biovitals® Sentinel, a remote patient management platform (Biofourmis, Boston MA, USA) that has been deployed in several countries across the world. It is designed to monitor COVID-19 patients remotely and longitudinally through a companion smartphone application and a comfortable armband wearable device. Patients enrolled into the COVID-19 clinics are given the option to record their cough sounds using the phone application. They are instructed to provide their spontaneous cough for at least three seconds, along with their self-reports of current symptoms and quality of life surveys. In this study, 321 spontaneous cough recordings collected from 112 confirmed COVID-19 positive patients have been included in our analysis.

Since a large portion of patients' self-recorded cough audios include mixed types of sounds, an ideal cough extraction model should be robust enough in differentiating cough from other sounds. Therefore, we supplemented our audio dataset with randomly selected non-cough audios from publicly available crowd-sourced datasets, including ESC-50 [10], DCASE2016 [11], Virufy [5], Coswara [12] and COUGHVID [13]. Given the limited sample size in our initial patient cohort, additional cough samples from the Virufy and Coswara datasets were also added to improve model performance and generalization. The number of cough and non-cough samples from external datasets were chosen to balance the final merged dataset.

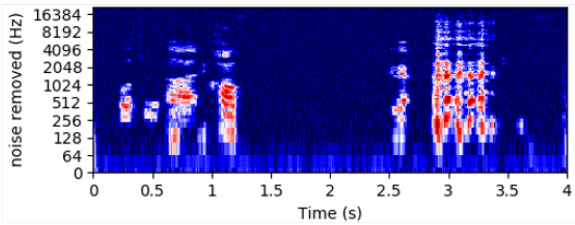
B. Audio Noise Reduction

Noise reduction is crucial in preparing the input audio records for subsequent prediction stages. Traditional noise filtering method requires prior knowledge and domain expertise in formulating appropriate filter designs to meet the signal output requirements. But a major issue with this filtering approach is that the noise spectrum profiles of cough recordings could differ greatly from sample to sample based on the background noise that is mixed with the file.

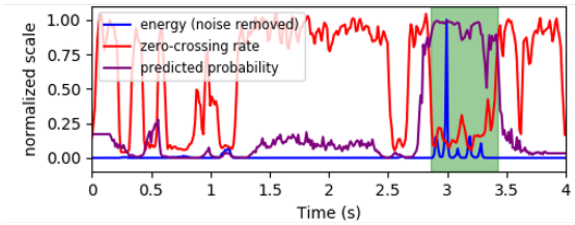
To solve this issue, we adopted the 'spectral gating' algorithm, an acoustic noise reduction technique based on the algorithm used in the Audacity(R) software [14]. This algorithm takes both an input audio sample and another audio clip with only background noise from the same or a similar waveform. It computes frequency band thresholds from the input noise clip to perform the filtering. However, this approach requires a noise sample to be manually



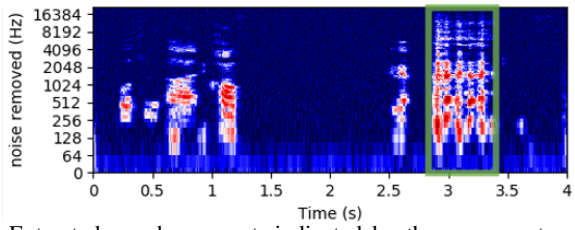
(a) Spectrogram representation of input audio sample.



(b) Spectrogram representation of noise-reduced audio.



(c) Normalized RMS energy, zero-crossing rate and CNN-predicted probability of noise-reduced audio vs. time. The colored window is generated by applying thresholds on these three measures.



(d) Extracted cough segment, indicated by the green rectangular box.

Fig. 1: Illustration of the high-level workflow of the noise reduction and cough localization algorithm, which extracts cough data from noisy audio samples and prepare them for the prediction stage.

extracted from each input file. To automate this extraction process in our work, we applied a threshold on short-time root-mean-square (RMS) energy of original audio clips to perform a preliminary foreground and background segmentation, as voiced segments tend to have much higher energy than ambient noise or other unvoiced audio segments [15]. The steady harmonic components were removed beforehand for better foreground/background separation, using the Librosa builtin 'decompose.hpss' function. The segmentation-generated noise samples were then fed into the spectral gating algorithm for noise reduction.

Figures 1a and 1b present an example audio spectrogram before and after applying the noise reduction algorithm respectively.

C. Manual Annotation For Outcome Generation

To train and validate the cough detection model and subsequently evaluate the temporal accuracy of our cough extraction algorithm, we require the onset and offset time points of all foreground segments being precisely labelled. Our Biovitals® Sentinel dataset was segmented and manually annotated by three researchers using a procedure as described below:

1) A threshold is applied on the normalized RMS energy of each frame window of length 23.2 millisecond (1024 samples at a 44.1 kHz sampling rate) in the noise-reduced audio files to extract all foreground segments. Two scorers independently label the onset and offset timestamps of each cough episode from foreground segments in each file. The remaining foreground segments are then assigned with 'non-cough' labels.
2) If the labels agree and the time points differ by less than 0.1 second, the averaged time points are taken as the final ground truth.
3) In case of a disagreement, a third annotator listens to the file, reviews the previous two versions of annotation and chooses to either select one annotation as ground truth, enter a new set of onset and offset times along with a cough label, or mark the file to be excluded if the label is difficult to determine.

The external datasets did not require any additional manual labelling, since those audio files are known to contain only one type of sound that has been already labelled.

The labelled segments are divided into train (n = 2337 total segments, 53% cough segments), validation (n = 226 total segment, 49% cough segments) and test set (n = 257 total segments, 52% cough segments) by an 80:10:10 ratio on the audio file level to avoid data leakages between sets.

D. Cough Detection CNN Classifier

After de-noising, annotating and splitting the audio segments, we then use them for developing the cough extraction algorithm. Our proposed algorithm is mainly built upon a cough detection classifier trained on short-time slices of audio segment samples. The input samples for this cough detection classifier are generated by transforming the manually extracted various-length audio segments into Mel-frequency cepstral coefficients (MFCCs) and then applying a sliding window of 0.3 second across the MFCC images with a step size of 1024 (23.2 millisecond) to extract short fixed-length input samples for training the cough detection model. We designed this 0.3-second sliding window width, which is narrower than that in most of other studies, in order to achieve better temporal precision.

The CNN structure is employed for the cough detection model, as it has been extensively used for audio signal processing in the literature [16], [17]. Our best-performing CNN structure is comprised of a single convolutional layer with 32 filters of size 5×3, a stride of $(2, 1)$, ReLU activation function, batch normalization and a 2×1 max-pooling layer with a 0.2 dropout rate. The learnt features are then flattened and passed into a fully connected layer of 20 neurons and ReLU activation function. The final layer is a 2-neuron sigmoid dense layer that generates binary predictions of 'cough' vs. 'non-cough' on input audio segments. All the CNN hyper-parameters were tuned based on the validation AUROC and in consideration of model over-fitting. The model was trained using the Adaptive gradient optimizer given its smooth converging effect for sparse image data.

The CNN model predicts a single probability for each 0.3-second audio slice. For model validation, the predicted probabilities for these 0.3-second slices are averaged into a single probability for every audio input segment.

E. Cough Onset And Offset Localization

The trained cough detection CNN model is used in conjunction with a signal processing algorithm to localize onsets and offsets of coughs from unknown audio clips. We obtain profiles of model predicted probability of cough, short-term energy and zero-crossing rate of the same audio by performing frame-level computation. Triple thresholding on normalized frame energy, zero-crossing rate and predicted probability is applied to extract cough segments. This hybrid approach is based on the fact that CNN predicted probability shows high discriminating power on cough vs. non-cough but lacks in temporal precision. Therefore, we leverage the good sensitivity and temporal precision of energy and zero-crossing rate measures to compensate the disadvantage of using the CNN model alone.

This algorithm workflow is illustrated in Fig. 1 using a real data example. The model-detected onsets and offsets of cough episodes are compared to the manually-annotated ones to validate the extraction performance. In this work we use the performance evaluation method described in [18] to compute episode-based and duration-based performance metrics. The episode-based metrics evaluate the proportion of true episodes that are correctly detected by the model in terms of episode counts, while the duration-based metrics evaluate the proportion in terms of duration in time. Note that if a model-extracted time point differs from the manual annotation by ≥ 0.1 second, the difference time window is considered as one false event and is counted towards performance calculation. Duration AUROC is derived by computing the duration true-positive rate (TPR) and false-positive rate (FPR) using different probability thresholds in the cough detection model while keeping the signal processing steps the same.

Silent episodes with very low energy or episodes shorter than 0.1 sec were excluded from the analysis and the subsequent COVID-19 diagnosis model.

F. COVID-19 CNN Classifier

The purpose of the COVID-19 diagnosis model is to distinguish COVID positive from COVID negative using patient provided cough samples, From all the cough audio samples used for the previous cough detection model, we excluded the ones with no COVID labels (positive or negative) for this model, leaving the dataset with 182 COVID positive

TABLE I: Performance metrics of the cough detection model for the test dataset.

	Sens.	PPV	Spec.	Acc.	F1	AUROC
Set threshold	0.955	0.934	0.927	0.942	0.945	0.987
Top performance	1.0	1.0	1.0	0.957	0.960	

and 181 COVID negative audio files. The COVID negative coughs are recorded from either healthy subjects or patients with other mainly respiratory diseases. We used repeated k-fold cross-validation instead of traditional train-test split to more accurately evaluate model performance given the limited dataset. The cross-validation split is performed on the patient level to prevent data leakage.

All the cough segments are first transformed into MFCCs and then fed into a CNN classifier. The same CNN architecture as in the cough detection model is re-used, with its final layer re-configured to predict the outcome labels of 'COVID positive' vs. 'COVID negative'.

In the test set, we generate a probability for every single 0.3-second slice of cough segments extracted from an audio file and then take the median of the predicted probabilities of all slices as the outcome to represent the likelihood of the file containing any cough event from a COVID-19 positive patient.

III. RESULTS

A. Cough detection and extraction

Prior to validation of the cough extraction algorithm, the cough detection model that uses manually extracted cough segments is first evaluated alone. The model performance metrics on the test set data are summarized in Table I. We present both the performance metrics using the set threshold that is selected to maximize the validation-set F1 score and the highest value for each single performance metric. The AUROC of our best-performing CNN model is 0.987. The corresponding accuracy is 94.2% with sensitivity of 95.5%, specificity of 92.7% and F1 score of 94.5%. The sensitivity, precision and specificity metrics all reach 100% when their optimal thresholds are used.

We then assessed the entire cough extraction system using the episode-based and duration-based performance metrics. The results are summarized in Table II. Our algorithm achieves a patient-average episode and duration F1 of 0.935 and 0.961 respectively and a duration AUROC of 0.986.

B. COVID-19 classification

To evaluate the COVID-19 diagnosis model, we conducted the 5-fold cross-validation procedure for 5 times and reported the mean and standard deviation of each performance metric from all the 25 runs in Table III.

IV. DISCUSSIONS

A. Cough detection and extraction

Our cough detection CNN model achieves a high level of accuracy as well as balanced performance in other metrics.

TABLE II: Episode and duration performance metrics of the cough extraction algorithm for the test dataset.

	Gross statistics	Patient-average statistics
Episode sensitivity	0.942	0.902
Episode PPV	0.958	0.971
Episode F1	0.950	0.935
Duration sensitivity	0.979	0.946
Duration specificity	0.997	0.995
Duration PPV	0.983	0.976
Duration NPV	0.997	0.976
Duration accuracy	0.995	0.977
Duration F1	0.981	0.961
Duration AUROC	0.993	0.986

TABLE III: Performance metrics of the COVID-19 classification model using 5-fold cross-validation repeated 5 times.

	Sens.	PPV	Spec.	Acc.	F1	AUROC
Mean (Std.)	0.953 (0.032)	0.970 (0.028)	0.958 (0.039)	0.958 (0.026)	0.961 (0.023)	0.981 (0.013)

It shows great potential to be tailored for different use cases since its highest sensitivity, precision and specificity are all equal to 100%. The fact that the model gives relatively lower episode-based sensitivity (0.902) than the duration-based sensitivity (0.946) is consistent with our observation that the model tends to generate slightly more false negatives for very short cough episodes (≤ 0.2 second), given that the model is trained using 0.3-second segments. However, since on the Biovitals® Sentinel platform patients are told to record audios for at least three seconds, these extremely short cough episodes only takes up a very small portion in each audio sample. It is reasonable to believe that these false negatives would not affect the overall system performance.

Most of previous works on cough detection use sensitivity and specificity as their main performance metrics. Based on our current knowledge, our model sensitivity and specificity outperforms the other audio-based cough detection systems in literature. It is noted that many published results have shown very high specificity (90.9% - 99.8%) at the cost of significantly lower sensitivity (70.5% - 88%) [7], [19]. Although the models introduced in [8] has shown an overall accuracy of 99.8%, its dataset only consists of night-time recordings of participants during their sleep, which include very little noise to cope with. Some other studies have achieved similar model performance only with the addition of other device data, such as contact microphones, respiratory inductance plethysmography, ECG sensors, and accelerometers [20]. Our proposed method, in contrast, shows high-level and balanced performance using only smartphone audio recordings in spite of a noisy and challenging dataset.

B. COVID-19 classification

Our COVID-19 identification CNN model has a mean AUROC of 0.981, demonstrating its predictive power to distinguish between COVID positive and COVID negative coughs with high accuracy. Besides, the model achieves

balanced performance with a mean of above 0.95 in each metric, which is comparable to the performance of COVID-19 classifiers in recent studies ([3]–[5], etc.) despite different datasets. It should be highlighted that our input data collection process only requires patient themselves to record audios using their smartphones, eliminating the need of in-person visits to hospitals or labs. In comparison, the data collection processes in previous works are controlled much more strictly, and additional steps of data pre-processing and cleaning that requires prior knowledge about the dataset were also performed. Many of these studies uses transfer learning techniques that incorporate pre-trained deep learning models like Resnet50; some have shown success in building ensemble models to further enhance the performance; others have proved that additional manually extracted acoustic features and patient meta-data could be useful as well. These results inspire and encourage us to keep optimizing our model for further performance improvements, given our promising preliminary results.

V. CONCLUSION

In this paper, we present a fully automated algorithm that extracts coughs from audio files and then screens subjects for COVID-19 infection. Our results show this algorithm is capable of handling low-quality audio recorded through a smartphone in an environment with various noises present. These noises not only include natural environmental sounds in audio background, but also refer to other foreground sounds with high amplitude, especially voiced sounds that resembles coughs.

Both the proposed cough extraction and the COVID-19 classification models give high level of accuracy, using the same CNN structure design of extremely low complexity. At the time of the research being conducted, we have only collected a limited number of audio samples; we could likely further improve the model performance if using a larger training dataset, incorporating patient meta-data (e.g. age, gender and medical history) and other manually extracted acoustic features, or building ensemble models by adding pre-trained deep learning models like ResNet50 into our proposed system. The proposed cascading system and CNN structure can also be potentially used for other cough analysis, patient health assessment and prediction tasks by slightly change the model setup and/or adding in other vitals and symptom data into the model which we have been already being collected through the Biovitals® Sentinel platform.

REFERENCES

[1] Worldometer. Coronavirus Update (Live): Cases and Deaths from COVID-19 Virus Pandemic, 2021.

[2] Lauren M. Kucirka, Stephen A. Lauer, Oliver Laeyendecker, Denali Boon, and Justin Lessler. Variation in False-Negative Rate of Reverse Transcriptase Polymerase Chain Reaction-Based SARS-CoV-2 Tests by Time Since Exposure, aug 2020.

[3] Jordi Laguarta, Ferran Hueto, and Brian Subirana. COVID-19 Artificial Intelligence Diagnosis Using Only Cough Recordings. *IEEE Open Journal of Engineering in Medicine and Biology*, 1:275–281, 2020.

[4] Ali Imran, Iryna Posokhova, Haneya N. Qureshi, Usama Masood, Muhammad Sajid Riaz, Kamran Ali, Charles N. John, MD Iftikhar Hussain, and Muhammad Nabeel. AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. *Informatics in Medicine Unlocked*, 20:100378, jan 2020.

[5] Ahmed Fakhry, Xinyi Jiang, Jaclyn Xiao, Gunvant Chaudhari, Asriel Han, and Amil Khanzada. Virufy: A Multi-Branch Deep Learning Network for Automated Detection of COVID-19. mar 2021.

[6] Madhurananda Pahar, Marisa Klopper, Robin Warren, and Thomas Niesler. COVID-19 Cough Classification using Machine Learning and Global Smartphone Recordings. *arXiv preprint arXiv:2012.01926*, 2020.

[7] Sandra Larson, Germán Comina, Robert H Gilman, Brian H Tracey, Marjory Bravard, and José W López. Validation of an automated cough detection algorithm for tracking recovery of pulmonary tuberculosis patients. *PloS one*, 7(10):e46229, 2012.

[8] Filipe Barata, Peter Tinschert, Frank Rassouli, Claudia Steurer-Stey, Elgar Fleisch, Milo Alan Puhan, Martin Brutsche, David Kotz, and Tobias Kowatsch. Automatic Recognition, Segmentation, and Sex Assignment of Nocturnal Asthmatic Coughs and Cough Epochs in Smartphone Audio Recordings: Observational Field Study. *Journal of medical Internet research*, 22(7):e18082, 2020.

[9] Michael Joseph Pettinati, Xiyu Zhang, Ali Jalali, Kuldeep Singh Rajput, and Nandakumar Selvaraj. Automatic and robust identification of spontaneous coughs from covid-19 patients. *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2021. submitted.

[10] Karol J Piczak. ESC: Dataset for environmental sound classification. In *Proceedings of the 23rd ACM international conference on Multimedia*, pages 1015–1018, 2015.

[11] Annamaria Mesaros, Toni Heittola, Emmanouil Benetos, Peter Foster, Mathieu Lagrange, Tuomas Virtanen, and Mark D Plumbley. Detection and classification of acoustic scenes and events: Outcome of the DCASE 2016 challenge. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 26(2):379–393, 2018.

[12] Neeraj Sharma, Prashant Krishnan, Rohit Kumar, Shreyas Ramoji, Srikanth Raj Chetupalli, R. Nirmala, Prasanta Kumar Ghosh, and Sriram Ganapathy. Coswara - A database of breathing, cough, and voice sounds for COVID-19 diagnosis. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2020-Octob:4811–4815, 2020.

[13] Lara Orlandic, Tomas Teijeiro, and David Atienza. The COUGHVID crowdsourcing dataset: A corpus for the study of large-scale cough analysis algorithms, sep 2020.

[14] Audacity Team. Audacity ® — Free, open source, cross-platform audio software for multi-track recording and editing., 2019.

[15] R. G. Bachu, S. Kopparthi, B. Adapa, and B. D. Barkana. Voiced/unvoiced decision for speech signals based on zero-crossing rate and energy. In *Advanced Techniques in Computing Sciences and Software Engineering*, pages 279–282. Springer Publishing Company, 2010.

[16] Kuniaki Noda, Yuki Yamaguchi, Kazuhiro Nakadai, Hiroshi G. Okuno, and Tetsuya Ogata. Audio-visual speech recognition using deep learning. *Applied Intelligence*, 42(4):722–737, jun 2015.

[17] Emre Cakir, Giambattista Parascandolo, Toni Heittola, Heikki Huttunen, and Tuomas Virtanen. Convolutional Recurrent Neural Networks for Polyphonic Sound Event Detection. *IEEE/ACM Transactions on Audio Speech and Language Processing*, 25(6):1291–1303, jun 2017.

[18] Prashanthan Sanders, Helmut Pürerfellner, Evgeny Pokushalov, Shantanu Sarkar, Marco Di Bacco, Bärbel Maus, and Lukas R.C. Dekker. Performance of a new atrial fibrillation detection algorithm in a miniaturized insertable cardiac monitor: Results from the Reveal LINQ Usability Study. *Heart Rhythm*, 13(7):1425–1430, jul 2016.

[19] Jesus Monge-Alvarez, Carlos Hoyos-Barcelo, Paul Lesso, and Pablo Casaseca-De-La-Higuera. Robust Detection of Audio-Cough Events Using Local Hu Moments. *IEEE Journal of Biomedical and Health Informatics*, 23(1):184–196, jan 2019.

[20] Thomas Drugman, Jerome Urbain, Nathalie Bauwens, Ricardo Chessini, Anne Sophie Aubriot, Patrick Lebecque, and Thierry Dutoit. Audio and contact microphones for cough detection. In *13th Annual Conference of the International Speech Communication Association 2012, INTERSPEECH 2012*, volume 2, pages 1302–1305, 2012.