

# Samsung R and D Bangalore DiCOVA 2021 Challenge System Report

Vishwanath Pratap Singh\*, Shashi Kumar\*, Ravi Shekhar Jha\*

Samsung Research and Development Institute Bangalore, India

vp.singh@samsung.com, sk.kumar@samsung.com, rs.jha@samsung.com

<sup>1</sup>

## Abstract

The COVID-19 pandemic has resulted in more than 125 million infections and more than 2.7 million casualties. In this report, we attempt to classify covid vs non-covid cough sounds using signal processing and deep learning methods. Air turbulence, vibration of tissues, movement of fluid through airways, opening and closure of glottis are some of the causes for production of acoustic sound signal during cough. However, Does the COVID-19 alter the acoustic characteristics of breathe, cough, and speech sounds produced through the respiratory system? This is an open question waiting for answers.

In this paper we incorporated novel data augmentation methods for cough sound augmentation, and multiple deep neural network architectures and methods along with handcrafted features. Our proposed system gives 14% absolute improvement in area under the curve(AUC).

**Index Terms:** COVID-19, acoustics, machine learning, respiratory diagnosis, healthcare

## 1. System Description

### 1.1. Methodology Overview

In first part of work, we incorporate corporate Vocal tract length perturbation(VTLP), spectrum interpolation, variational autoencoder(VAE), noise based data augmentation methods to increase the totaltrain data size. Then, we explore the various multiple deep neural network architectures Long short-term memory(LSTM), Convolutional neural network(CNN), residual neural network(ResNet) and its variants.

### 1.2. Pre-processing

First the raw cough speech signal is passed to speech activity detector and then normalized between -1 to 1. Then, set of handcrafted feature along with 39 dimensional mel-frequency cepstral coefficient(MFCC) are extracted.

### 1.3. Feature Description

We have experimented with 39 dimensional mfcc and 24 dimensional handcrafted features to capture the time domain and frequency domain variability between COVID-19 negative and positive cough sounds. Handcrafted feature includes per frame energy of the signal(1D), fundamental frequency(1D), first four formants(4D), alpha ratio(1D) with cutoff frequency at 1400 Hz, relative average perturbation(1D), spectral flatness(1 D), kurtosis(1D), spectral contrast(7D), second order spectral polynomial(2D), spectral centroid(1D), spectral rolloff(1D), and spectral bandwidth(1D), root mean square value of the signal(1D), and zero crossing rete of the signal(1D). Then delta

<sup>1</sup>\* Represents the equal contribution

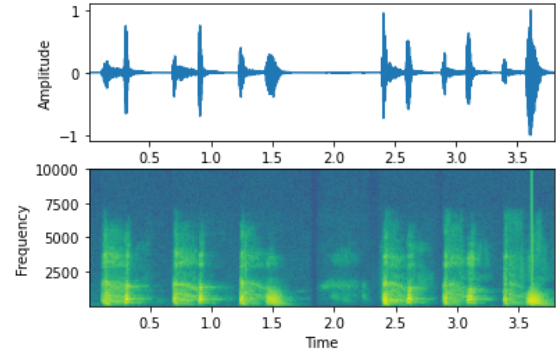


Figure 1: COVID-19 positive cough sound signal and spectrogram.

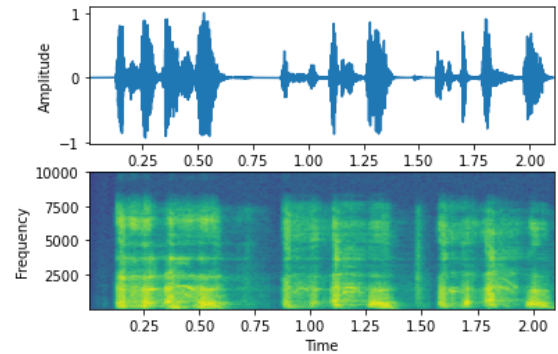


Figure 2: Non-covid cough sound signal and spectrogram. It is noted that higher proportion of signal energy lies in higher frequency range compare COVID-19 cough sound signal.

and double delta of the feature vector is computed to form 169 (3\*63) dimensional feature.

### 1.4. Classifier Description

We have experimented with standard machine learning algorithms such as, support vector machine(SVM) and random forest classifiers, and deep neural network architecture such as, LSTM, CNN and ResNet. We experiment with number of layers, sequence length and class weights in cross entropy(CE) loss.

### 1.5. Results

In initial submissions, we observed significant drop (around 5%) in AUC from validation sets to blind test set which may arise due to mismatch in speaker variabilities and acoustic environment. To gauge the performance of our classifiers effec-

Table 1: results of different models on combined validation and augmented validation set

model	fold	#layers	hidden dim	type	wtPos	norm	seq length	weight decay	AUC	specificity
LSTM	1	1	48	uni	1	utt-wise	30	0	75.33	67.5
LSTM	1	1	48	uni	1	utt-wise	30	0.001	76.7	62.11
LSTM	2	1	64	uni	1	utt-wise	30	0	76.48	59.5
LSTM	2	1	64	uni	3	utt-wise	30	0	75.78	61
LSTM	1	1	64	uni	1	utt-wise	40	0	76.84	55
LSTM	2	1	48	uni	2	utt-wise	30	0	74.97	52.17
LSTM	1	1	48	uni	2	utt-wise	30	0	77.08	59
LSTM	1	1	48	uni	2	utt-wise	40	0	76.52	55.9
LSTM	5	1	48	uni	1	utt-wise	30	0	73.64	50.31
LSTM	1	1	128	uni	1	utt-wise	30	0	79	56.52
LSTM	2	2	48	uni	1	utt-wise	30	0	75.03	55.28
LSTM	3	2	64	uni	1	utt-wise	30	0	77.17	60.87
LSTM	1	1	64	uni	2	utt-wise	40	0.001	75.27	57.14
LSTM	3	1	48	uni+5-fold	2	global	40	0	75.47	54.04
LSTM	2	1	48	seq-to-concat	1	utt-wise	30	0	74.58	55.28
LSTM	4	1	48	seq-to-concat	1	utt-wise	30	0	74.89	48.45
LSTM	2	1	48	seq-to-concat	1	utt-wise	30	0.001	72.02	55.9
LSTM	3	1	48	seq-to-last1	7	utt-wise	30	0.001	74.98	52.17
LSTM	5	1	48	seq-to-last1	2	global	30	0	74.36	60.24
LSTM	4	1	48	uni	3	utt-wise	20	0	76.63	50.31
LSTM	1	1	48	uni	1	global	30	0	75.03	46.58
LSTM	1	1	48	uni	2	global	40	0	76.58	63.35
LSTM	4	1	48	uni	2	global	40	0	76.99	57.8
LSTM	1	1	48	uni	2	global	50	0	76.01	51.55
LSTM	1	1	48	bidir	2	global	40	0	76.07	63.35
LSTM	3	1	48	bidir	2	global	30	0	76.11	62.11
LSTM	1	1	48	uni+coughvid	2	utt-wise	30	0	76.72	67.08
LSTM	1	1	32	uni+AUROC loss	1	global	30	0	77.54	69.57
LSTM	2	1	32	uni+AUROC loss	1	global	30	0	76.13	61.49
ResNet-18	2	NA	32	filter=3	1	utt-wise	inCh=3	0	75.96	53.42
ResNet-34	3	NA	32	filter=3	1	utt-wise	inCh=3	0	76.13	54.66
CNN	1	4	64	filter=5	2	utt-wise	inCh=3	0	75.96	59.63
CNN	1	4	64	filter=5	1	utt-wise	inCh=3	0	75.76	49.69
CNN	3	2	128	filter=5	1	utt-wise	inCh=3	0	75.46	56.28
ensemble results on our set									<b>85.89</b>	<b>70.19</b>
ensemble results on blind test set (leaderboard)									<b>83.93</b>	<b>70.83</b>

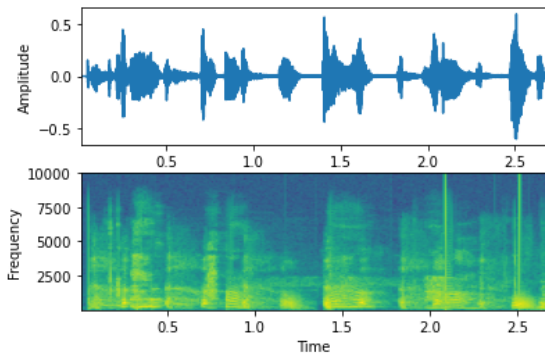


Figure 3: Spectrum interpolation based augmentation of COVID-19 positive cough sound signal and spectrogram shown in Figure 1.

spective augmented versions. We use VTLP augmentation for speaker variabilities and additive noise augmentation for acoustic mismatch.

tively, we augment validation set of each fold and then calculate AUC on the default validation set for each fold as well as on re-