



Robust Cough Feature Extraction and Classification Method for COVID-19 Cough Detection Based on Vocalization Characteristics

Xueshuai Zhang^{1,2,§}, Jiakun Shen^{1,2,§}, Jun Zhou^{1,2}, Pengyuan Zhang^{1,2}, Yonghong Yan^{1,2,†}, Zhihua Huang³, Yanfen Tang⁴, Yu Wang⁴, Fujie Zhang⁴, Shaoxing Zhang⁵, Aijun Sun⁶

¹Key Laboratory of Speech Acoustics and Content Understanding, Institute of Acoustics, Chinese Academy of Sciences, China

²University of Chinese Academy of Sciences, China

³School of Information Science and Engineering, Xinjiang University, China

⁴Beijing Ditan Hospital Capital Medical University, China

⁵Peking University Third Hospital, China

⁶Dalian Public Health Clinical Center, China

{zhangxueshuai, shenjiakun, yanyonghong}@hcccl.ioa.ac.cn

Abstract

A fast, efficient and accurate detection method of COVID-19 remains a critical challenge. Many cough-based COVID-19 detection researches have shown competitive results through artificial intelligence. However, the lack of analysis on vocalization characteristics of cough sounds limits the further improvement of detection performance. In this paper, we propose two novel acoustic features of cough sounds and a convolutional neural network structure for COVID-19 detection. First, a time-frequency differential feature is proposed to characterize dynamic information of cough sounds in time and frequency domain. Then, an energy ratio feature is proposed to calculate the energy difference caused by the phonation characteristics in different cough phases. Finally, a convolutional neural network with two parallel branches which is pre-trained on a large amount of unlabeled cough data is proposed for classification. Experiment results show that our proposed method achieves state-of-the-art performance on Coswara dataset for COVID-19 detection. The results on an external clinical dataset Virufy also show the better generalization ability of our proposed method.

Index Terms: cough, acoustic features, convolutional neural networks, pre-trained model, COVID-19 detection

1. Introduction

Since the outbreak of COVID-19 in December 2019, over 440 million confirmed cases and 5.9 million deaths have been reported by the WHO. The variants of the original virus are more infectious and are spreading rapidly. One of the effective measures to control the epidemic is to screen out COVID-19 individuals from the crowd.

At present, the most commonly used method for COVID-19 detection is Reverse Transcription Polymerase Chain Reaction (RT-PCR). However, this method is restricted to time, resources, and the way of detection with lower sensitivity in reality [1]. Other diagnostic methods, such as CT scans, have strict requirements on medical equipment [2, 3]. Due to these limitations, a repeatable, real-time, cost-effective and individually achievable method is of vital importance to detect COVID-19 individuals in large-scale populations effectively.

As cough is one of the main symptoms of COVID-19 [4], many researchers have used cough sounds to distinguish COVID-19 individuals from healthy individuals. Some of these studies used machine learning methods for classification with handcrafted features as inputs [5, 6, 7, 8]. Short-Time Fourier Transform (STFT) based features are extracted and input into deep neural networks in some other researches [9, 10, 11]. As pre-trained model has achieved excellent performance in other tasks, it has also been applied to cough-based COVID-19 detection [12, 13, 14, 15]. The above studies use acoustic features that have already been widely used in other audio tasks. However, few studies utilize the characteristics of cough itself. In our previous work [16], we proposed a multi-band neural network to detect COVID-19 individuals by using the characteristic that cough energy is concentrated in low frequency [17].

In this paper, we further propose two energy-related features to characterize cough sounds. A single cough consists of explosive, intermediate and voiced phases [18]. Since cough is short and intense compared to speech, the energy is more variable at different phases and frequencies. To describe the dynamic information in adjacent frames and frequency bands, we propose a time-frequency differential feature (TFDF). We also propose an energy ratio feature to calculate the energy span of different phases of cough. Considering the limited amount of cough sounds from COVID-19 individuals, we also use pre-trained models for parameter initialization. Unlike previous researches that used pre-trained models on ImageNet [19] or Audio Set [20], we use contrastive learning strategy to learn the latent features from large amounts of unlabeled cough sounds. As a result, our proposed method achieves an improvement on AUC by 4.28% and 10.47% absolutely on Coswara [21] and Virufy [22] respectively. In short, the main contributions of this paper are summarized as follows:

- We propose a time-frequency differential feature to describe the dynamic information of cough sounds in adjacent frames and frequency bands.
- We propose an energy ratio feature to measure the energy span of different phases of cough.
- We propose a convolutional neural network structure with two parallel branches and use pre-trained model based on a large number of unlabeled cough data for parameter initialization.

[§]Equal contribution

[†]Corresponding author

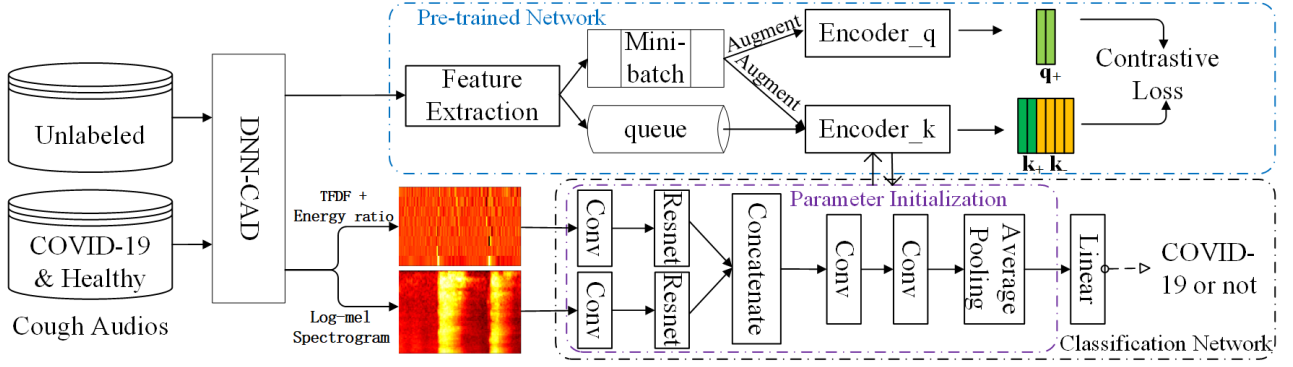


Figure 1: The proposed framework.

2. Method

The proposed framework is shown in Figure 1. We first use cough and non-cough recordings to train an independent deep neural network based cough activity detection module (DNN-CAD) [23] to remove the silence and speech segments in cough recordings. Log-mel spectrograms and proposed features are then extracted and fed into two parallel branches of the classification network to give the probability of cough belonging to COVID-19 individuals. Apart from the final linear layer, the parameters of the classification network are initialized by the weights of the pre-trained model.

2.1. Time-Frequency Differential Feature (TFDF)

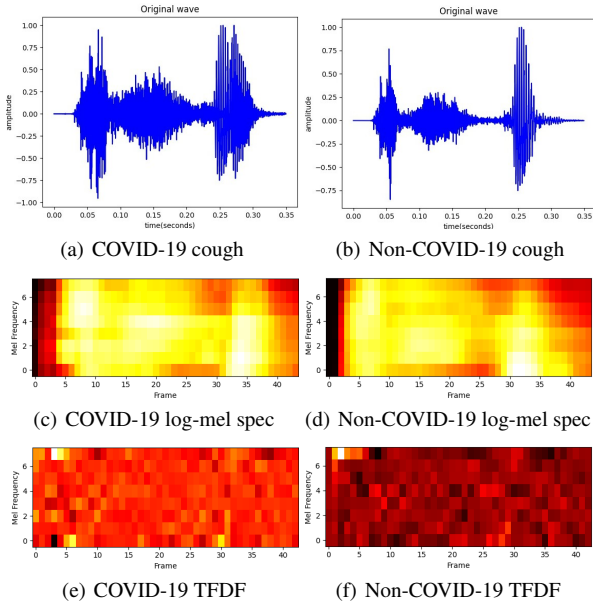


Figure 2: Feature analysis for COVID-19 and Non-COVID-19 cough.

The TFDF is generated by calculating the energy difference at frequency and time domain of the log-mel spectrograms. Since the cough energy is concentrated in low frequency and varies in different frequency bands, we first use the energy difference between adjacent frequency bands to describe the dynamic changes in frequency domain. Considering the instan-

taneity of cough sounds in time domain, we then calculate the energy difference at frame level. This calculation process is shown in Equations 1 - 3:

$$\Delta E_F(t, m) = E(t, m) - E(t, m + 1) \quad (1)$$

$$\Delta E_T(t, m) = E(t + 1, m) - E(t, m) \quad (2)$$

$$\begin{aligned} \Delta E_{FT}(t, m) &= \Delta E_F(t + 1, m) - \Delta E_F(t, m) \\ &= \Delta E_T(t, m) - \Delta E_T(t, m + 1) \\ &= E(t + 1, m) - E(t, m) - \\ &\quad E(t + 1, m + 1) + E(t, m + 1) \end{aligned} \quad (3)$$

where t denotes the frame number and m denotes the sub-band number. $E(t, m)$ means the energy at the t th frame and m th sub-band of the log-mel spectrogram. $\Delta E_F(t, m)$, $\Delta E_T(t, m)$, $\Delta E_{FT}(t, m)$ represent the energy difference in frequency domain, time domain, both time and frequency domain respectively, in which ΔE_{FT} is TFDF. We select the cough sounds with similar waveforms from COVID-19 and healthy individuals. Figure 2(a) and Figure 2(b) show the waveforms. The energy distributions are similar in the log-mel spectrograms with eight mel filters shown in Figure 2(c) and Figure 2(d). From the TFDF shown in Figure 2(e) and Figure 2(f), a more obvious difference in energy distributions can be observed.

Algorithm 1 Calculating energy ratio

Input: Waveform: $wav[n]$

Hyper-parameters: $win_len, hop_len, threshold, \rho$

Output: Energy ratio: $ratio$

$framenum = (n - win_len) / hop_len + 1$

for $i \leftarrow 1$ to $framenum$ **do** //calculate the energy list e

for $j \leftarrow 1$ to win_len **do**

$e[j] \leftarrow e[j] + wav[(i - 1) * hop_len + j]^2$

end for

$e[i] \leftarrow \sqrt{e[i]}$

end for

$e \leftarrow \text{SortDescend}(e)$ //sort e in descending order

$e \leftarrow \text{Remove}(e, threshold)$ // remove item less than threshold

$length \leftarrow \text{len}(e)$ // get the length of the list

for $i \leftarrow 1$ to $length * \rho$ **do**

$max_energy \leftarrow max_energy + e[i]$

$min_energy \leftarrow min_energy + e[length - i]$

end for

$ratio \leftarrow min_energy / max_energy$

2.2. Energy Ratio

There are differences in peak energy and energy distribution in different cough phases among different types of cough [18]. Therefore, we propose an energy ratio feature by calculating the ratio of minimum energy to maximum energy of cough’s different phases to measure the energy span of cough sounds. The calculation steps are illustrated in algorithm 1.

The *win.len* and *hop.len* are the window length and hop length when framing. *Threshold* is average energy of silence segments and ρ is the percent of the maximum and minimum energy, which is 0.2 in our experiments. Figure 3 shows the distribution of energy ratios of cough sounds in Coswara dataset. The average energy ratio of cough sounds in healthy people is 0.035, which is higher than 0.03 in COVID-19 individuals.

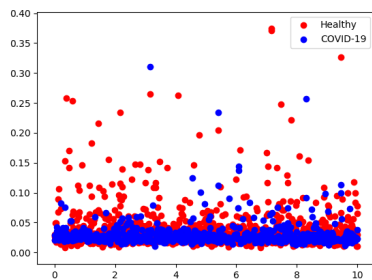


Figure 3: Distribution of energy ratios

2.3. Pre-trained Model on Cough Sounds

The proposed convolutional neural network structure is shown in Figure 1. In classification network, the ResNet block consists of four building blocks of ResNet18 [24].

To avoid the influence of irrelevant information caused by non-cough sounds during pre-training model, we use a large number of unlabeled cough sounds to learn the latent features through contrastive learning proposed in [25]. After DNN-CAD, mini-batches and a queue are constructed and log-mel spectrograms and proposed features are extracted. The features in mini-batches are augmented in different ways to form positive sample pair and the features in queue are considered as negative samples. These features are input into two encoders and the parameters of these two encoders are updated by minimizing the contrastive loss and momentum update, respectively. The specific implementation refers to [25].

3. Experiment

3.1. Datasets

Three open-sourced cough sound datasets are used for pre-training without any label information. The data distribution is shown in Table 1.

Then Coswara dataset is used for classification, which is one of the most widely used open-sourced crowd-sourced cough datasets to distinguish COVID-19 individuals from healthy individuals. Each person records a heavy cough and a shallow cough in this dataset. We randomly divide the data into train, validation and test set in a ratio of 3:1:1. The cough recordings of the same person only appear in one set to ensure speaker independence. In addition, we also use an external clinical dataset from Virufy to test the generalization ability of the proposed model. The data distribution is shown in Table 2.

Table 1: Datasets for pre-trained model

Dataset	samples	total time(after CAD)
ESC50 [26]	40	1min8s
IAtos [27]	4657	3h50min43s
COUGHVID [28]	8996	9h35min53s

Table 2: Datasets for classification

Dataset	Set	COVID-19	Healthy
Coswara [21]	train	531	1188
	validation	175	395
	test	177	396
Virufy [22]	test	48	73

3.2. Experiment setup

All the audios are first resampled to 16kHz. We use a variable chunk size in training to adapt to different audio lengths in test set, which means the audio will be repeated or cropped until the audio length is equal to chunk size.

The log-mel spectrograms are extracted with a window length of 512 samples, hop length of 128 samples and 128 mel filters. Except for the number of mel filters, the parameters of TFDF are the same as those of log-mel spectrograms.

When pre-training, we apply data augmentation on the fly, including adding noise, time shift, time stretch, and SpecAugment [29]. When fine-tuning the classification network, we also use SpecAugment to overcome overfitting. Model parameters are initialized by default when the pre-trained model is not used.

The mini-batch size is 64 and we use Adam optimizer with a weight decay of 1e-3. The learning rate is 1e-4 and multiplied by 0.7 every 20 epochs.

Area under receiver operating characteristic (ROC-AUC) is used to measure the model performance, which is calculated by sensitivity and specificity at different probability thresholds.

3.3. Results and Analysis

The experimental results are shown in Table 3. We list the results of some other studies on Coswara dataset on the top of Table 3. In experiment 1, the 128-dimensional log-mel spectrograms are input into one of the branches of the classification network as baseline, which achieved an AUC of 0.8163 on the test set. Based on our previous research [16], we also split the 128-dimensional log-mel spectrograms into low and high parts along frequency axis and input them into the two branches of the classification network in experiment 2.

3.3.1. Performance of proposed methods

In experiment 3, 128-dimensional log-mel spectrograms and TFDF with 8 mel filters are input into the parallel branches of the classification network. The AUC increases by 2.03% absolutely compared to experiment 1. Although experiment 2 achieves better performance than experiment 1 by separating log-mel spectrograms into low and high frequency bands, the AUC of experiment 3 has still been improved by 1.13% compared to that of experiment 2 by adding TFDF under the condition of the same model parameter size. This result indicates that TFDF can capture similar dynamic changes in cough sounds of

Table 3: Experimental results on Coswara. ExId means experiment id. Sen means sensitivity. Spe means specificity. LMS means log-mel spectrogram. ER means energy ratio. PM means using pre-trained model for parameter initialization.

ExId	Feature	AUC	Sen	Spe
	Mohammed[30]	0.77	0.71	0.82
	Van [31]	0.8121	0.4833	0.9513
1	LMS	0.8163	0.5480	0.8889
2	LMS(split)	0.8253	0.6102	0.8737
3	Ex 1 + TFDF	0.8366	0.5311	0.9369
4	Ex 3 + ER	0.8422	0.5876	0.9116
5	Ex 4 + PM	0.8591	0.5989	0.9116

the same groups in time and frequency domain, which is helpful for distinguishing COVID-19 individuals from healthy people.

In experiment 4, the AUC increases by 0.56% compared to experiment 3 after the energy ratio is concatenated to TFDF. Although the improvement is quite slight, the result shows that the energy ratio represents a certain characteristic of cough sounds. A plausible explanation is that TFDF describes the dynamic information in adjacent frames and frequency bands, while energy ratio further supplements the energy span information between different cough phases.

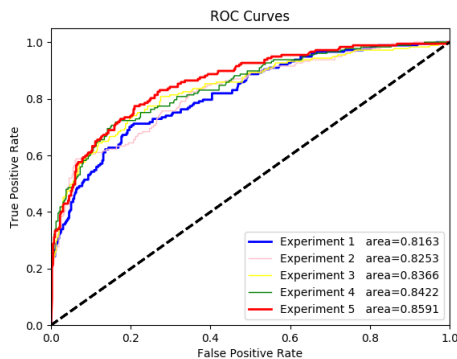


Figure 4: ROC Curves

In experiment 5, model parameters are initialized by the weights of pre-trained model. With the same input features as experiment 4, the AUC increases by 1.69%. Finally, our proposed methods are superior to the model in experiment 1 on AUC with an improvement of 4.28% absolutely. Figure 4 shows the ROC curves of these experiments.

3.3.2. Influence of TFDF parameters on performance

The number of mel filters in TFDF is obtained through the following experiments. We use 8, 16, 32, 64 and 128 mel filters to extract TFDF with different dimensions. The experimental results show that the classification performance is quite similar when using 8, 16, 32 mel filters, but the performance will decrease when the number of mel filters increases to 64 and 128. The reason may be that higher dimensional TFDF pays more attention to the uniqueness of a sample rather than the similarity between samples of the same class, which results in overfitting on train set and performance degradation on test set.

We also explore the influence of differential scale on classification performance when extracting TFDF. To compare with

Table 4: Influence of TFDF parameters on performance. T, F, T-F mean calculating the energy difference in time, frequency, both time and frequency dimension, respectively.

scale	Mel filters	AUC	Sensitivity	Specificity
T-F	8	0.8366	0.5311	0.9369
T-F	16	0.8360	0.5198	0.9495
T-F	32	0.8366	0.5254	0.9394
T-F	64	0.8345	0.4972	0.9444
T-F	128	0.8108	0.4802	0.9343
T	8	0.8317	0.5424	0.9242
F	8	0.8360	0.5480	0.9192

experiment 3, the differential feature is obtained by calculating energy difference only along time axis or frequency axis. In Table 4, the model with either time differential feature or frequency differential feature outperforms the model in experiment 1, and the latter achieves a higher AUC. The reason may be that the difference in time domain may be affected by the times of cough, while the difference in frequency domain can better reflect the impact of COVID-19 on vocalization characteristics.

3.3.3. The results on Virufy test set

The model in experiment 1 achieves an AUC of 0.6376 on Virufy test set, which is lower than the AUC on Coswara. This may be due to the large difference in data distribution between the datasets. After adding TFDF and energy ratio features, the model in experiment 4 achieves an AUC of 0.7423, which shows a better generalization ability of our proposed model.

Table 5: Results on Virufy test set.

ExId	AUC	Sen	Spe
1	0.6376	0.3750	0.7123
4	0.7423	0.5833	0.7397

4. Conclusions

In this paper, a cough-based deep learning framework has been proposed to detect COVID-19 individuals in the crowd. According to the characteristics of cough sounds, we propose a time-frequency differential feature to describe the dynamic information and an energy ratio feature to measure the energy span of different phases of the cough. By using model pre-trained on a large amount of cough sounds for parameter initialization, our model achieves an improvement of 4.28% absolutely on AUC on Coswara dataset. The 10.47% increase in AUC on Virufy test set also shows the better generalization ability of our model.

For future work, we will further investigate the classification performance on COVID-19 individuals with different symptoms. Other sounds such as speech and breath will be combined to improve classification performance.

5. Acknowledgements

This work is partially supported by Beijing Municipal Science and Technology Project (No.Z211100002521020) and Hygiene and Health Development Scientific Research Fostering Plan of Haidian District Beijing (HP2021-11-50102).

6. References

- [1] R. Liu, H. Han, F. Liu, Z. Lv, K. Wu, Y. Liu, Y. Feng, and C. Zhu, "Positive rate of rt-pcr detection of sars-cov-2 infection in 4880 cases from one hospital in wuhan, china, from jan to feb 2020," *Clinica chimica acta*, vol. 505, pp. 172–175, 2020.
- [2] T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, Q. Tao, Z. Sun, and L. Xia, "Correlation of chest ct and rt-pcr testing for coronavirus disease 2019 (covid-19) in china: a report of 1014 cases," *Radiology*, vol. 296, no. 2, pp. E32–E40, 2020.
- [3] G. Wang, X. Liu, J. Shen, C. Wang, Z. Li, L. Ye, X. Wu, T. Chen, K. Wang, X. Zhang *et al.*, "A deep-learning pipeline for the diagnosis and discrimination of viral, non-viral and covid-19 pneumonia from chest x-ray images," *Nature biomedical engineering*, vol. 5, no. 6, pp. 509–521, 2021.
- [4] C. Gomes, "Report of the who-china joint mission on coronavirus disease 2019 (covid-19)," *Brazilian Journal of Implantology and Health Sciences*, vol. 2, no. 3, 2020.
- [5] C. Brown, J. Chauhan, A. Grammenos, J. Han, A. Hasthanasombat, D. Spathis, T. Xia, P. Cicuta, and C. Mascolo, "Exploring automatic diagnosis of covid-19 from crowdsourced respiratory sound data," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020, pp. 3474–3484.
- [6] M. Pahar, M. Klopper, R. Warren, and T. Niesler, "Covid-19 detection in cough, breath and speech using deep transfer learning and bottleneck features," *Computers in biology and medicine*, vol. 141, p. 105153, 2022.
- [7] F. Avila, A. H. Poorjam, D. Mittal, C. Dognin, A. Muguli, R. Kumar, S. R. Chetupalli, S. Ganapathy, and M. Singh, "Investigating feature selection and explainability for covid-19 diagnostics from cough sounds," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, vol. 6. International Speech Communication Association, 2021, pp. 4246–4250.
- [8] S. Bhosale, U. Tiwari, R. Chakraborty, and S. K. Kopparapu, "Contrastive learning of cough descriptors for automatic covid-19 preliminary diagnosis," in *Proc. Interspeech*, vol. 2021, 2021, pp. 946–50.
- [9] A. Mallol-Ragolta, H. Cuesta, E. Gómez, and B. W. Schuller, "Cough-based covid-19 detection with contextual attention convolutional neural networks and gender information," in *22nd Annual Conference of the International Speech Communication Association, INTERSPEECH 2021*, 2021, pp. 4236–4240.
- [10] A. Imran, I. Posokhova, H. N. Qureshi, U. Masood, M. S. Riaz, K. Ali, C. N. John, M. I. Hussain, and M. Nabeel, "Ai4covid-19: Ai enabled preliminary diagnosis for covid-19 from cough samples via an app," *Informatics in Medicine Unlocked*, vol. 20, p. 100378, 2020.
- [11] K. K. Lella and A. Pja, "Automatic covid-19 disease diagnosis using 1d convolutional neural network and augmentation with human respiratory sound based on parameters: cough, breath, and voice," *AIMS Public Health*, vol. 8, no. 2, p. 240, 2021.
- [12] J. Laguarda, F. Hueto, and B. Subirana, "Covid-19 artificial intelligence diagnosis using only cough recordings," *IEEE Open Journal of Engineering in Medicine and Biology*, vol. 1, pp. 275–281, 2020.
- [13] J. Han, T. Xia, D. Spathis, E. Bondareva, C. Brown, J. Chauhan, T. Dang, A. Grammenos, A. Hasthanasombat, A. Floto *et al.*, "Sounds of covid-19: exploring realistic performance of audio-based digital testing," *NPJ digital medicine*, vol. 5, no. 1, pp. 1–9, 2022.
- [14] P. Bagad, A. Dalmia, J. Doshi, A. Nagrani, P. Bhamare, A. Mahale, S. Rane, N. Agarwal, and R. Panicker, "Cough against covid: Evidence of covid-19 signature in cough sounds," *arXiv preprint arXiv:2009.08790*, 2020.
- [15] J. Harvill, Y. R. Wani, M. Hasegawa-Johnson, N. Ahuja, D. Beiser, and D. Chestek, "Classification of covid-19 from cough using autoregressive predictive coding pretraining and spectral data augmentation," in *22nd Annual Conference of the International Speech Communication Association, INTERSPEECH 2021*. International Speech Communication Association, 2021, pp. 4261–4265.
- [16] J. Shen, X. Zhang, W. Wang, Z. Huang, P. Zhang, and Y. Yan, "Cough-based covid-19 detection with multi-band long-short term memory and convolutional neural networks," in *Proceedings of the 2nd International Symposium on Artificial Intelligence for Medicine Sciences*, 2021, pp. 209–215.
- [17] J. Knocikova, J. Korpas, M. Vrabec, and M. Javorka, "Wavelet analysis of voluntary cough sound in patients with respiratory diseases," *J Physiol Pharmacol*, vol. 59, no. Suppl 6, pp. 331–40, 2008.
- [18] H. Chatzarrin, A. Arcelus, R. Goubran, and F. Knoefel, "Feature extraction for the differentiation of dry and wet cough sounds," in *2011 IEEE international symposium on medical measurements and applications*. IEEE, 2011, pp. 162–166.
- [19] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, no. 3, pp. 211–252, 2015.
- [20] J. F. Gemmeke, D. P. Ellis, D. Freedman, A. Jansen, W. Lawrence, R. C. Moore, M. Plakal, and M. Ritter, "Audio set: An ontology and human-labeled dataset for audio events," in *2017 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2017, pp. 776–780.
- [21] N. Sharma, P. Krishnan, R. Kumar, S. Ramoji, S. Chetupalli, R. Nirmala, P. Kumar Ghosh, and S. Ganapathy, "Coswara-a database of breathing, cough, and voice sounds for covid-19 diagnosis," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, vol. 2020. International Speech Communication Association, 2020, pp. 4811–4815.
- [22] G. Chaudhari, X. Jiang, A. Fakhry, A. Han, J. Xiao, S. Shen, and A. Khanzada, "Virufy: Global applicability of crowdsourced and clinical datasets for ai detection of covid-19 from cough," *arXiv preprint arXiv:2011.13320*, 2020.
- [23] J. Amoh and K. Odame, "Deep neural networks for identifying cough sounds," *IEEE transactions on biomedical circuits and systems*, vol. 10, no. 5, pp. 1003–1011, 2016.
- [24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [25] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 9729–9738.
- [26] K. J. Piczak, "Esc: Dataset for environmental sound classification," in *Proceedings of the 23rd ACM international conference on Multimedia*, 2015, pp. 1015–1018.
- [27] D. T. Pizzo, S. Esteban, and M. Scetta, "Iatos: Ai-powered pre-screening tool for covid-19 from cough audio samples," *arXiv preprint arXiv:2104.13247*, 2021.
- [28] L. Orlandic, T. Teijeiro, and D. Atienza, "The coughvid crowdsourcing dataset, a corpus for the study of large-scale cough analysis algorithms," *Scientific Data*, vol. 8, no. 1, pp. 1–10, 2021.
- [29] D. S. Park, W. Chan, Y. Zhang, C.-C. Chiu, B. Zoph, E. D. Cubuk, and Q. V. Le, "SpecAugment: A simple data augmentation method for automatic speech recognition," *Proc. Interspeech 2019*, pp. 2613–2617, 2019.
- [30] E. A. Mohammed, M. Keyhani, A. Sanati-Nezhad, S. H. Hejazi, and B. H. Far, "An ensemble learning approach to digital corona virus preliminary screening from cough sounds," *Scientific Reports*, vol. 11, no. 1, pp. 1–11, 2021.
- [31] H. Van Truong and L. Pham, "A cough-based deep learning framework for detecting covid-19," *arXiv preprint arXiv:2110.03251*, 2021.