2022 IEEE Healthcare Innovations and Point of Care Technologies (HI-POCT)
March 10-11, 2022. Texas A&M University - EnMed, Houston, TX, USA

2022 IEEE Healthcare Innovations and Point of Care Technologies (HI-POCT) | 978-1-6654-9615-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/HI-POCT54491.2022.9744064

CoughNet-V2: A Scalable Multimodal DNN Framework for Point-of-Care Edge Devices to Detect Symptomatic COVID-19 Cough

Hasib-Al Rashid[1], Mohammad M. Sajadi[2] and Tinoosh Mohsenin[1]

Abstract— **With the emergence of COVID-19 pandemic, new attention has been given to different acoustic bio-markers of the respiratory disorders. Deep Neural Network (DNN) has become very popular with the audio classification task due to its impressive performance for speech detection, audio event classification etc. This paper presents CoughNet-V2 - a scalable multimodal DNN framework to detect symptomatic COVID-19 cough. The framework was designed to be implemented on point-of-care edge devices to help the doctors at pre-screening stage for COVID-19 detection. A crowd-sourced multimodal data resource which contains subjects' cough audio along with other relevant medical information was used to design the CoughNet-V2 framework. CoughNet-V2 shows multimodal integration of cough audio along with medical records improves the classification performance than that of any unimodal frameworks. Proposed CoughNet-V2 achieved an area-under-curve (AUC) of 88.9% for the binary classification task of symptomatic COVID-19 cough detection. Finally, measurement of the deployment attributes of the CoughNet-V2 model onto processing components of an NVIDIA TX2 development board is presented as a proposition to bring the healthcare system to consumers' fingertips.**

Clinical relevance— **CoughNet-V2 will help medical practitioners to asses whether the patients need intensive medical help without physically interacting with them.**

Index Terms— **COVID-19, Cough Sound Classification, Multimodal Deep Neural Networks, Point-of-Care Edge Devices.**

I. INTRODUCTION

The COVID-19 pandemic has wreaked havoc on people's health, as well as their social and economic lives. It damages a variety of body structures and organs. At the onset of pandemic, it is important to detect the COVID-19 patients and separate them from other lungs disease patients to give them better attention before the medical test confirms their COVID-19 infection. Most individuals are unconcerned about their breathing and respiratory health, and they miss the fact that their lungs are vital organs that can be infected or damaged. Because the symptoms of respiratory disorders are usually interchangeable, this might lead to misdiagnosis. Therefore, developing a diagnostic discriminant is crucial for identifying a quick and correct diagnosis of respiratory symptoms and taking the appropriate steps. For the treatment of respiratory disorders, it is vital to make an accurate and timely diagnosis. COVID-19 characterized by a variety of symptoms including a dry cough, a fever, exhaustion and dyspnea (shortness of breath) etc. and correspond significantly with different races, genders, and age groups at different phases of the disease's development. Fever was reported by approximately 70% of COVID-19 verified patients in conjunction with dry cough [1]. In contrast to the elderly, who are the most affected group, clinical case studies show that the young population is less prone to have COVID-19-related symptoms [2].

Coughing is a typical symptom of respiratory problems [3]. Analyzing the cough sound during treatment can provide valuable information about the coughing pathophysiological actions that contribute to specific cough patterns. [4]. Cough sound changes are thought to be an important indicator of the severity of a respiratory infection and the effectiveness of treatment [4].

In the domain of medical research, machine learning and deep learning are very popular for medical computer vision and medical image classification tasks [5]–[7]. Deep Neural Networks also shows promising performance in the task of audio event and signal classification [8], [9]. Our previous research has shown that respiratory sounds can be used to detect a variety of respiratory disorders [10]. Moreover, we have shown with the help of deep learning algorithms, we could detect cough sounds out of different other environmental sounds [11]. Recently, multimodal deep learning gained much attention as the overall model performance is improved when different modalities of data is considered [12]–[14].

This paper introduces CoughNet-V2, a scalable and multimodal DNN model running on point-of-care edge devices to evaluate patients utilizing passively collected cough audio, and self-entry information (such as age, gender and fever). The proposed CoughNet-V2 framework has the ability to make a significant effect by bringing preventative healthcare to users' fingertips and estimating the need for them to visit clinics and have themselves further evaluated using more specialized test-kits or facilities. The main contributions of this work include:

- Analyze the open-sourced COVID-19 cough audio dataset statistically to justify extracting a sufficiently balanced dataset out of it.
- Propose CoughNet-V2, a scalable multimodal DNN on the extracted COVID-19 cough audio dataset in conjunction with the medical information reported by

[1]Hasib-Al Rashid and Tinoosh Mohsenin are with Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County, USA. (hrashid1, tinoosh)@umbc.edu
[2]Mohammad M. Sajadi is with Institute of Human Virology, School of Medicine, University of Maryland, USA. msajadi@ihv.umaryland.edu

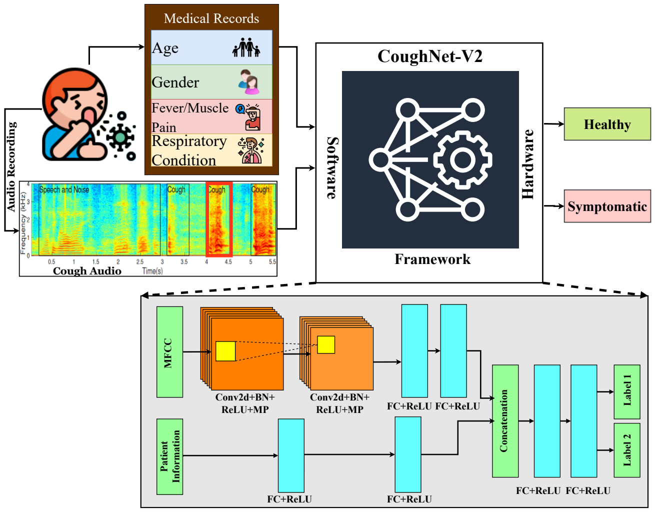


Fig. 1: The high-level overview and the detailed model architecture of proposed CoughNet-V2. The multimodal inputs of the model are MFCC 2d image of size (101, 40) which is converted from audio recordings with sampling rate of 44.1 KHz and patients' self-reported medical records. Here, Conv2d = 2 dimensional convolutional neural networks, BN = Batch Normalization, ReLU = Rectified Linear Unit, MP = Max Pooling, and FC = Fully Connected Layer

the patients.

- Train an unimodal deep convolutional neural network on the extracted COVID-19 cough audio dataset without considering the medical information and compare the unimodal and multimodal DNN approach in terms of model performance.
- Implement the CoughNet-V2 to the TX2 embedded system and evaluate its CPU and GPU implementation characteristics.

II. Related Works

As diagnosis COVID-19 signature from cough audio has become an active area of research to artificial intelligence community, a bunch of unimodal and multimodal COVID-19 audio dataset have been presented [15]–[20]. Out of them only [17], [18] have been made publicly available for the researcher community. [16], [20] provides their dataset upon signing their user agreement. Authors in [21]–[23] proposed unimodal machine learning and deep learning based approach to classify whether the cough sound has COVID-19 signature or not. Authors in [24]–[26] proposed multimodal machine learning and deep learning approaches to classify COVID cough audios. However, authors in [24], [26] had to manually select different features out of COVID-19 cough sound whereas our proposed CoughNet-V2 automatically selects features from the mel frequency cepstral coefficients (MFCCs) derived from the cough audios. Mel-frequency cepstrum (MFC) is the short-term power power spectrum representation of a sound, based on the linear cosine transform of a log power spectrum on a nonlinear mel-scale of the frequency. There are MFCC (mel-frequency cepstral coefficients) that form an MFC. A form of cepstral representation is used to create these sounds. To have more accurate representation of human hearing, the mel-frequency cepstrum is the way to go, as the frequency bands are evenly spaced on this scale, which is closer to how our ears actually respond. Authors in [26] presented similar approach to ours. However, they have worked on Coswara [18] dataset whereas we have worked on COUGHVID dataset [27]. To date, COUGHVID [27] is the largest dataset that has been openly published which leads to better classification performance. Moreover, we have shown the implementation of the CoughNet-V2 to the TX2 embedded system and evaluate its CPU and GPU implementation characteristics which shows acceptability of our proposed CoughNet-V2 for point-of-care edge devices.

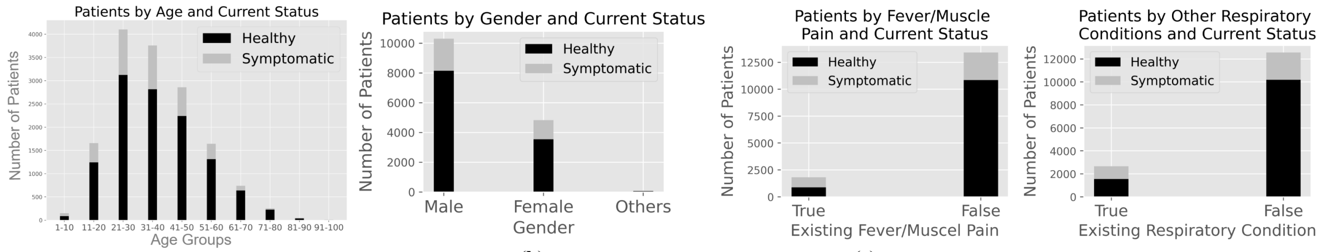
TABLE I: Details of the network architecture for processing MFCCs of the cough audio.

Layers	Description	Output
Input Layer	Audio MFCC Vector	499 × 13
Conv2D	Kernels = 16 × (3 × 3) - BN - ReLU	499 × 13 × 16
MaxPooling2D	Pool size = (2×2), 20% Dropout	249 × 6 × 16
Conv2D	Kernels = 32 × (3 × 3) - BN - ReLU	249 × 6 × 32
MaxPooling2D	Pool size = (2×2), 20% Dropout	124 × 3 × 32
Flatten	124 × 3 × 32	11904
Dense	Neurons = 64 - ReLU - 20% Dropout	64
Dense	Neurons = 32 - ReLU - 20% Dropout	32

III. CoughNet-V2 Framework

The high level overview of the proposed CoughNet-V2 framework and the more detailed architecture of the framework is presented in Figure 1. CoughNet-V2 can take an audio recordings of the cough and self-reported medical information from the user. As the input is in the form of audio recordings, we converted the audio recordings into MFCC 2D images, where rows correspond to features from MFCCs and columns correspond to time (window). The data is then separated into window frames in order to extract characteristics, as the right windows are critical for distinguishing between static and continuous signals. Windowing involves first standardizing the independent variables and then creating sliding T windows. Then the window frames are forwarded to the two dimensional CNN layers followed by fully connected layes. The detailed layer parameters are presented in Table I. Simultaneously, the medical records vectors are passed by two fully connected layers having 64 and 32 nodes. Then the output of the upper network (cough audio processing) and the lower networks (medical records processing) are concatenated and then forwarded to two fully connected layers having 256 and 128 nodes respectively. At the end, the output for the binary classification is seen in the form of the probability distribution of the last fully connected layer with the Softmax activation function. Additionally, other performance metrics i.e AUC, F1-score, Precision, Sensitivity and Specificity were measured. We trained our model with binary cross-entropy loss and SGD optimizer with 0.6 as momentum.

For the unimodal training experiment, we passed the output of the fully connected layer to the binary classification layer and measured the performance metrics.



(a) (b) (c) (d)

Fig. 2: Statistics of the COUGHVID database that contains both cough audio samples and self reported medical information from 15218 subjects. Orinially the dataset has 3 class labels but due to absence of original RT-PCR test reports, we considered both Symptomatic and COVID-19 classes as Symptomatic class. a) Breakdown of each class with respect to 10 age groups, b) Breakdown of each class with respect to gender, c) Break-down of each class with respect to reported fever/muscle pain, d) Break-down of each class with respect to other respiratory conditions

IV. Experimental Results and Analysis

In this section, CoughNet-V2 is evaluated with in-depth analysis using COUGHVID [17] datasets for COVID signature detection in cough sound along with the respective experimental results.

A. *Data Sets*

We evaluated CoughNet-V2 for COVID-19 signature detection in cough sound on the publicly available dataset, COUGHVID [17]. Although this dataset has more than 25,000 cough audio recordings, only 15,218 subjects' data contains both cough audio samples and self reported medical information. Moreover, some participants did not reported their medical information while uploading their cough audios. The dataset has some silent audio files which we excluded based on the cough detection accuracy which was mentioned in the dataset. Originally the dataset had 3 class labels, Healthy, Symptomatic and COVID-19 but due to absence of original RT-PCR test reports, we considered both Symptomatic and Covid-19 classes as Symptomatic class. After all those considerations, we got total 2527 symptomatic patients' data and 2663 healthy subjects' data, total 5190 participants' data were selected for this experiments. Then we selected 3 seconds of the time frame and randomly upsample our dataset to 10,000 audio files where 5241 number of samples were for healthy class and 4756 number of samples were for symptomatic class. We divided 10000 data sample into train-test sets with selecting 20% of the original data for test sets (8000 training and 2000 test data).

What motivates us to use this particular dataset for our multimodal CoughNet-V2 framework is that it contains some specific information (such as subjects' age, gender, previous respiratory conditions, existing fever/muscle pain etc.) which can be co-related with the corresponding subjects' cough audio so that those information can contribute towards improvement of model performance. One of our previous study [10] shows inclusion of age groups as knowledge vector into the respiratory symptoms classification model can improve the model accuracy upto 5%.

TABLE II: CoughNet-V2 model performance metrics across different modalities for detecting COVID-19 signatures in cough sounds

CoughNet-V2	Unimodal	Multimodal	Improvement
AUC (%)	83.4	88.9	5.5
F-1 Score (%)	81.1	86.7	5.5
Precision (%)	81.7	86.3	4.6
Sensitivity (%)	82.0	85.6	3.6
Specificity (%)	82.3	85.8	3.5

B. *Results*

Table II shows five standard evaluation metrics (AUC, F-1 Score, Precision, Specificity, Sensitivity/Recall) to evaluate the CoughNet-V2 framework on test dataset for multimodal experiments. We also presented the result without the related information from the patient, termed it as unimodal case, and evaluated under similar performance metrics for unimodal experiments. From comparing the results, it can be said that CoughNet-V2 significantly improves its performance for multimodal experiments as combination of both audio and medical information data complement each other.

V. Commercial Off-The-Shelf Device Deployment

The CoughNet-V2 is designed to be flexibly deployable for point-of-care edge devices where the proposed deep learning models trained on the CoughNet-V2 can be deployed on various edge processing engines. At least two hardware-level features are attributed to all DNN models: model size and number of computations per inference, both of which are upper-bounded by the platform resources they deploy on, or by the inference deadline. Both the hardware resource restrictions and the assessment latency should meet the application goals when putting all of the framework's components together. The trained CoughNet-V2 model is deployed on two mobile CPUs, including Denver (dual-core) and ARM-Cortex A57 (quad-core), along with an embedded CPU+GPU implementation with varying frequency settings, after setting the batch-size to 1. The TX2 development board provides precise on-board power metering for all of the settings. The implementation is summarized in Table III, from which we can see the

TABLE III: Implementation of the CoughNet-V2 model to commercial off-the-shelf devices including a dual-core Denver CPU, a quad-core ARM A57 CPU, and a combination of ARM CPU + Pascal GPU from the NVIDIA TX2 board.

Configuration	CPU Freq. (MHz)	GPU Freq. (MHz)	Power (mW)	Latency (S)	Performance (GFLOP/S)	Energy (J)	Energy Eff (GFLOPS/W)
Denver CPU	345	-	763	13.0	0.005	9.91	0.005
	2035	-	2753	0.8	0.092	2.20	0.033
ARM A57 CPU	345	-	1056	4.6	0.015	4.85	0.014
	2035	-	3472	0.6	0.123	2.08	0.035
TX2 CPU+GPU	2035	1300.5	8892	0.1	0.742	0.88	0.083

least power dissipating implementation (Denver with a low frequency) takes 13 seconds to classify one frame, whereas the most energy-efficient implementation (CPU+GPU) dissipates approximately 11× more power to classify the same frame in 0.1 seconds.

VI. CONCLUSIONS

We presented CoughNet-V2, a scalable multimodal DNN framework that employs as much correlated information as a dataset provides in an attempt to exploit deep learning algorithms to detect the signature of COVID-19 into cough sounds. We have showed that by combining both auditory and supplementary information for a selection of reasonably balanced dataset out of a publicly released COUGHVID database, the detection AUC of the trained model increases by 5.5%. Finally, we test our CoughNet-V2 model on a dual-core Denver CPU, a quad-core ARM Cortex A57 CPU, and a heterogeneous CPU+GPU implementation from the NVIDIA TX2 development board to see how they perform when deployed to a point-of-care edge device.

ACKNOWLEDGMENT

We acknowledge the support of the University of Maryland, Baltimore, Institute for Clinical Translational Research (ICTR) and the National Center for Advancing Translational Sciences (NCATS) Clinical Translational Science Award (CTSA) grant number UL1TR003098.

REFERENCES

[1] X. Zhao *et al.*, "Incidence, clinical characteristics and prognostic factor of patients with covid-19: a systematic review and meta-analysis," *MedRxiv*, 2020.

[2] P.-I. Lee, Y.-L. Hu, P.-Y. Chen, Y.-C. Huang, and P.-R. Hsueh, "Are children less susceptible to covid-19?" *Journal of Microbiology, Immunology, and Infection*, 2020.

[3] J. Korpáš and Z. Tomori, *Cough and Other Respiratory Reflexes./Kašel' a Iné Respiračné Reflexy*. Veda, 1979.

[4] J. Korpáš, J. Sadloňová, and M. Vrabec, "Analysis of the cough sound: an overview," *Pulmonary pharmacology*, vol. 9, no. 5-6, pp. 261–268, 1996.

[5] H.-A. Rashid *et al.*, "A low-power lstm processor for multi-channel brain eeg artifact detection," in *2020 21th International Symposium on Quality Electronic Design (ISQED)*. IEEE, 2020.

[6] M. Khatwani, H.-A. Rashid *et al.*, "A flexible multichannel eeg artifact identification processor using depthwise-separable convolutional neural networks," *ACM Journal on Emerging Technologies in Computing Systems (JETC)*, vol. 17, no. 2, pp. 1–21, 2021.

[7] A. N. Mazumder, H.-A. Rashid, and T. Mohsenin, "An Energy-Efficient low power LSTM processor for human activity monitoring," in *2020 IEEE 33rd International System-on-Chip Conference (SOCC) (SOCC 2020)*, Sep. 2020, in press.

[8] S. Hershey *et al.*, "Cnn architectures for large-scale audio classification," in *2017 ieee international conference on acoustics, speech and signal processing (icassp)*. IEEE, 2017, pp. 131–135.

[9] H. Lee *et al.*, "Unsupervised feature learning for audio classification using convolutional deep belief networks," *Advances in neural information processing systems*, vol. 22, pp. 1096–1104, 2009.

[10] M.Hosseini *et al.*, "Neural networks for pulmonary disease diagnosis using auditory and demographic information," in *epiDAMIK 2020: 3rd epiDAMIK ACM SIGKDD International Workshop on Epidemiology meets Data Mining and Knowledge Discovery*. ACM, 2020, pp. 1–5, in press.

[11] H.-A. Rashid *et al.*, "Coughnet: A flexible low power cnn-lstm processor for cough sound detection," in *2021 IEEE 3rd International Conference on Artificial Intelligence Circuits and Systems (AICAS)*. IEEE, 2021, pp. 1–4.

[12] P. R. Ovi *et al.*, "Aris: A real time edge computed accident risk inference system," in *2021 IEEE International Conference on Smart Computing (SMARTCOMP)*, 2021, pp. 47–54.

[13] A. Sarkar and M. Rahnemoonfar, "Vqa-aid: Visual question answering for post-disaster damage assessment and analysis," in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 2021, pp. 8660–8663.

[14] Z. Guo *et al.*, "Deep learning-based image segmentation on multimodal medical imaging," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 3, no. 2, pp. 162–169, 2019.

[15] P. Bagad *et al.*, "Cough against covid: Evidence of covid-19 signature in cough sounds," *arXiv preprint arXiv:2009.08790*, 2020.

[16] C. Brown *et al.*, "Exploring automatic diagnosis of covid-19 from crowdsourced respiratory sound data," *arXiv preprint arXiv:2006.05919*, 2020.

[17] L. Orlandic *et al.*, "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.

[18] N. Sharma *et al.*, "Coswara–a database of breathing, cough, and voice sounds for covid-19 diagnosis," *arXiv preprint arXiv:2005.10548*.

[19] G. Chaudhari *et al.*, "Virufy: Global applicability of crowdsourced and clinical datasets for ai detection of covid-19 from cough," *arXiv preprint arXiv:2011.13320*, 2020.

[20] M. Cohen-McFarlane *et al.*, "Novel coronavirus cough database: Nococoda," *IEEE Access*, vol. 8, pp. 154 087–154 094, 2020.

[21] V. Despotovic *et al.*, "Detection of covid-19 from voice, cough and breathing patterns: Dataset and preliminary results," *Computers in Biology and Medicine*, p. 104944, 2021.

[22] M. Z. Hossain *et al.*, "Covidenvelope: A fast automated approach to diagnose covid-19 from cough signals," *medRxiv*, 2021.

[23] P. Suppakitjanusant *et al.*, "Identifying individuals with recent covid-19 through voice classification using deep learning," *Scientific Reports*, vol. 11, no. 1, pp. 1–7, 2021.

[24] A. Ponomarchuk *et al.*, "Project achoo: A practical model and application for covid-19 detection from recordings of breath, voice, and cough," *arXiv preprint arXiv:2107.10716*, 2021.

[25] A. Pal and M. Sankarasubbu, "Pay attention to the cough: Early diagnosis of covid-19 using interpretable symptoms embeddings with cough sound signal processing," in *Proceedings of the 36th Annual ACM Symposium on Applied Computing*, 2021, pp. 620–628.

[26] S. R. Chetupalli *et al.*, "Multi-modal point-of-care diagnostics for covid-19 based on acoustics and symptoms," *arXiv preprint arXiv:2106.00639*, 2021.

[27] L. Orlandic, T. Teijeiro, and D. Atienza, "The coughvid crowdsourcing dataset: A corpus for the study of large-scale cough analysis algorithms," *arXiv preprint arXiv:2009.11644*, 2020.