IBM IEEE CAS/EDS
AI Compute Symposium 2021
October 13-14

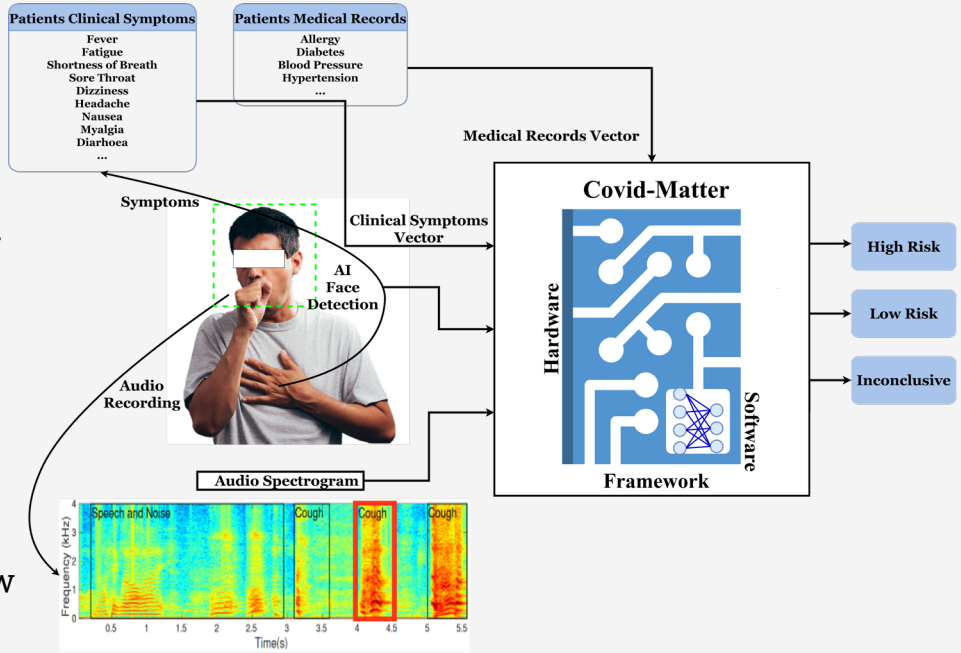
COVID-Matter: A Scalable Multimodal Sensory Machine Learning Framework for Severity Detection of Respiratory Diseases and Pandemic Prevention

Hasib-Al Rashid
Ph. D. Student
University of Maryland, Baltimore County

Introduction and Motivation

UMBC

- COVID-Matter is a novel software-hardware framework for collecting multi-modal data and applying modern machine learning methods to extract a variety of kinds of information tailored specifically to assess the risk progression of COVID-19.
- COVID-Matter framework can be extended for any kind of respiratory problem related solution i.e. COPD, URI etc. problems as well.
- COVID-Matter can potentially provide early assessment for anyone at anywhere to let them know whether they need medical attention.



Research Goals

UMBC

Stage 1

- Collect and analyze human-related respiratory sounds as well as symptom datasets using public resources

Stage 2

- Modify existing deep learning and language models for respiratory, audio, visual, and symptom datasets to meet analytical performance metrics of sensitivity, specificity, reliability and accuracy.

Stage 3

- Explore the most correlated group of sensory data in the models.
- Examine accuracy, specificity, sensitivity, computation complexity and speed of detection for different models with respect to different combinations of sensory data.

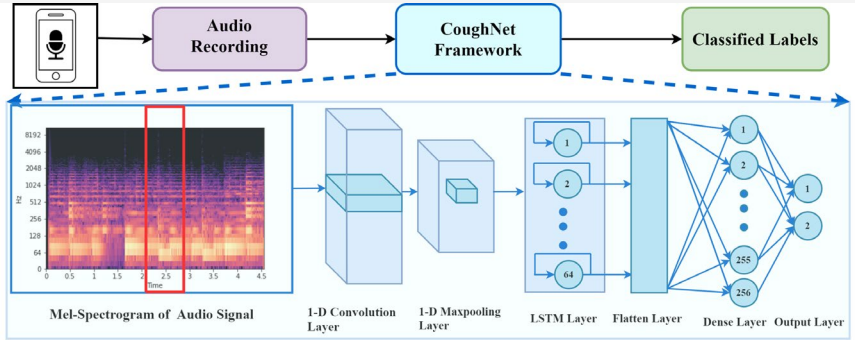
Stage 4

- Benchmark against a triage accuracy level if existing and analyze the accuracy, latency, computational and system performance overheads on a variety of heterogeneous embedded hardware architectures that include CPUs and GPUs.

Case Study 1: Cough Detection

UMBC

- Proposed CoughNet, a flexible software hardware CNN-LSTM framework that can take audio recordings and be configured for detecting cough sounds in it.
  - Perform input audio window size tuning, network architecture optimization with the goal of reducing computation complexity and memory size for resource constrained hardware implementation while maintaining the accuracy requirements.



ESC-50 [3]:

- 2,000 audio recordings of normal environmental sounds.
- 50 equally distributed classes including “coughing”, with each class having 40 audio recordings.
- All the audio recordings are 5 seconds in length
- Stored as single-channel audio waveform files at 44.1 kHz sampling rate.

FSDKaggle2018 [4]:

- 11,073 audio recordings of acoustic Scenes and Events sounds.
- 41 different classes including “coughing”
- Stored as 16bit PCM coded audio waveform files at 44.1 kHz sampling rate.

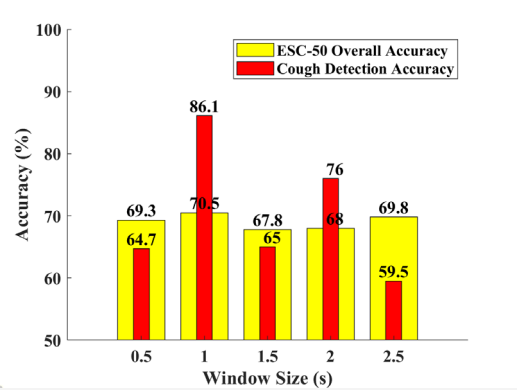
Coughvid [5]:

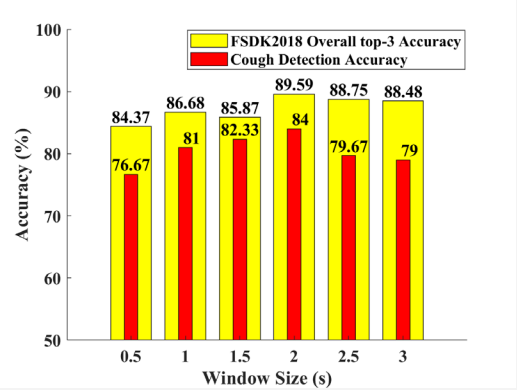
- Crowdsourced, validated, and publicly-available dataset of cough recordings.
- Provides over 20,000 cough recordings donated by participants.
- The COUGHVID dataset contributes over 820 expert-labeled coughs, from 4 different experts.
- All the audio recordings are 10 seconds in length

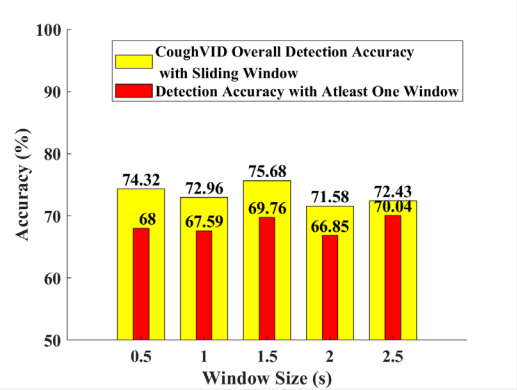
[3] K. J. Piczak, “Esc: Dataset for environmental sound classification,” in Proceedings of the 23rd ACM international conference on Multimedia, 2015, pp. 1015–1018.
[4] E. Fonseca et al., “General-purpose tagging of freesound audio with audioset labels: Task description, dataset, and baseline,” arXiv preprint arXiv:1807.09902, 2018.
[5] 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.

Case Study 1: Cough Detection

Software Results and Analysis







ESC-50 [3]:

- As for the performance on cough detection, 1s window shows good and balanced performance of extracting distinctive feature so it is chosen for our implementation scenario.

FSDKaggle2018 [4]:

- As for the performance on cough detection, 2s window shows good and balanced performance of extracting distinctive feature so it is chosen for our implementation scenario.

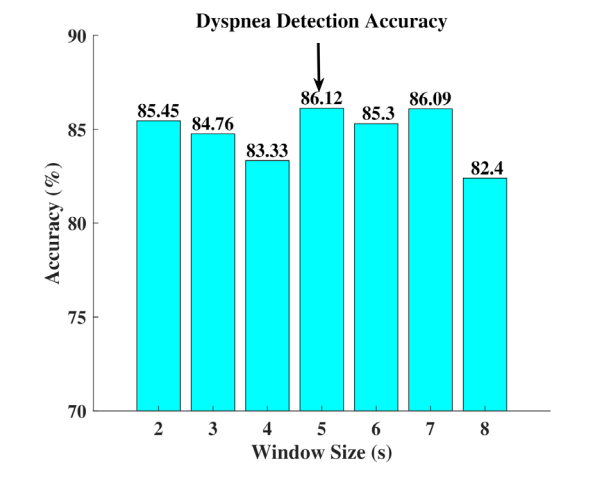
Coughvid [5]:

- As for the performance on cough detection, 1.5s window shows good and balanced performance of extracting distinctive feature so it is chosen for our implementation scenario.

Case Study 2: Dyspnea Detection

UMBC

- We proposed RespiratorNet for dyspnea detection, with a dataset collected from our participants.
- For each participant, we record two audio recordings. One is the sound of reading an article paragraph normally, and the other one is reading the same paragraph after some strenuous exercises, so that some gasp sounds would be included. We label the two audio recordings as normal and dyspnea accordingly.
- The recordings are recorded by various devices and then re-sampled at a sampling rate of 44.1~kHz. Each recording has a length between 30 to 60 seconds.
- After window extraction with different configurations, we could have about 3000 windows to be divided into train, validation, and test sets, while making sure that no window in the test set is overlapped with any window in the train set.
- Most of the model configurations are the same as the previous work. One difference is that we do not apply silence filtering for this case study, since audio recordings may include gasps. The other one is that we use window-wise prediction at testing, since we are doing a binary classification on the relatively small dataset.



Case Study 3: Detection of Respiratory Sound with Demographic Information

UMBC

- We evaluate the performance of RespiratorNet only with auditory input. In this one, we also include demographic information.
- The dataset (ICBHI 2017 challenge) we use for this case study comprises 920 recordings collected from 126 participants with annotations unequally disperse among 8 forms of respiratory conditions, including Upper Respiratory Tract Infection (URTI), Asthma, Chronic Obstructive Pulmonary Disease (COPD), Lower Respiratory Tract Infection (LRTI), Bronchiectasis, Pneumonia, and Bronchiolitis. The length of each recording varies from 10 to 90 seconds, often be controlled with 20 second samples.
- While the majority of this dataset are COPD-diagnosed participants, by taking only audio recordings captured by Welch Allyn Meditron Master Elite Electronic Stethoscope, one of the four devices used for this dataset, we generate a random subset encompassing 63 participants. We split it into a semi-balanced train and a test set of 52 and 11 participants that include 5 types of pulmonary classes. In consequence, we eliminate Asthma, Pneumonia, and LRTI.
- We performed a series of experiments, from audio input only, to merging the age group information with auditory signal. Table presents the two sets of studies, suggesting that the COPD and healthy conditions are diagnosed with higher accuracy and resulting in a total test accuracy increased by 5% when the demographic information is considered.

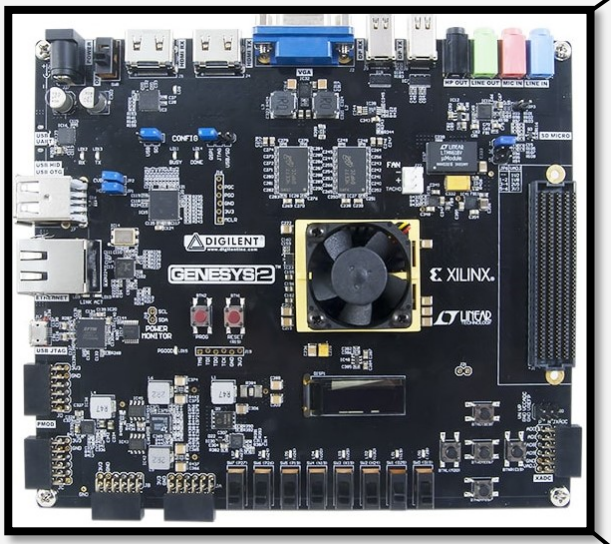
Table 1. Respiratory sound classification accuracy and model complexity with and without taking the demographic information into account.

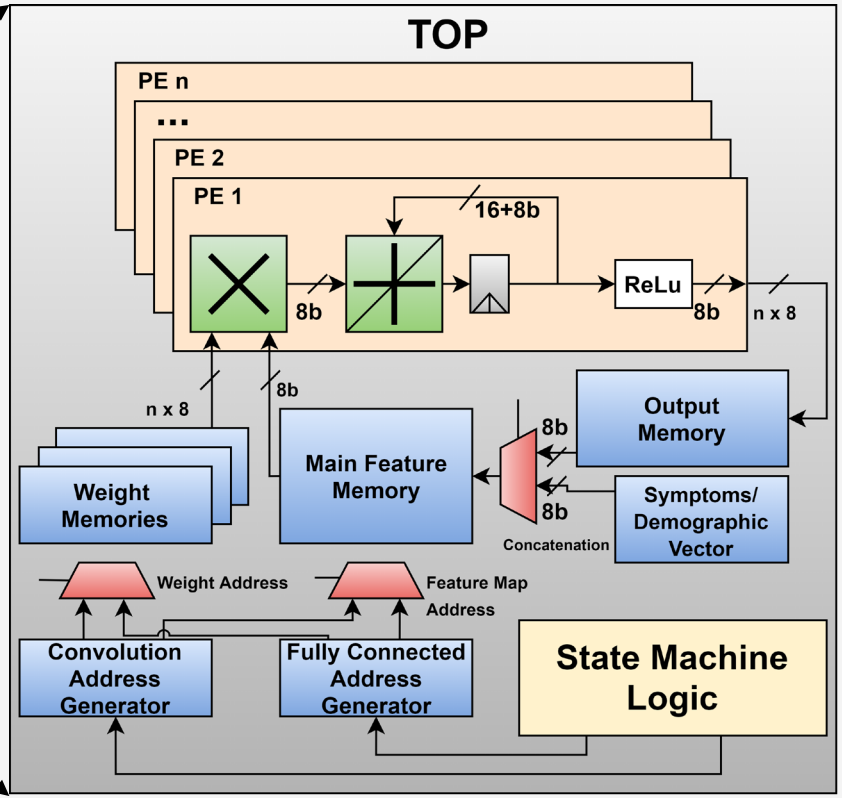
DCNN characteristics	Sensitivity (%)					Accuracy (%)
	URTI	Healthy	COPD	Bronchiec.	Bronchiol.	Test
Without Demographic Info.	21	66	96	88	4	78
With Demographic Info.	16	72	100	88	15	**83 (+5%)**

Hardware Implementation

UMBC

Kintex-7 FPGA Board





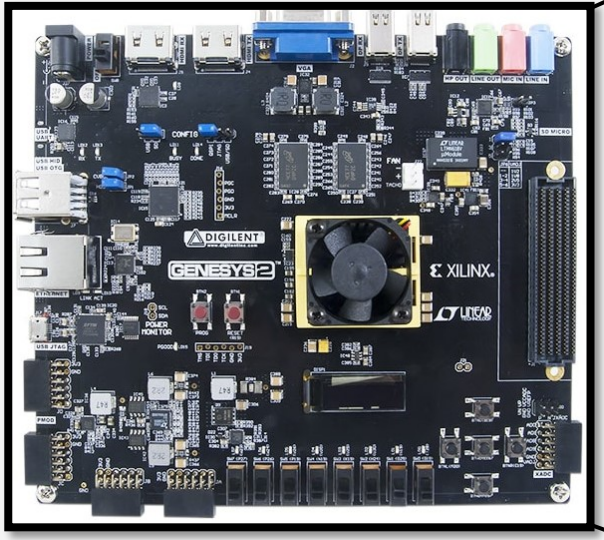
Primary implementation requirements for the hardware RTL (Register-Transfer Level) configuration:

- low utilization overhead,
- low power consumption,
- infrequent memory accesses.

Hardware Implementation

UMBC

Kintex-7 FPGA Board



Architecture	This work			
Application	Cough Detection (FSDKaggle2018)	Dyspnea Detection	Respiratory Sounds with demographic info	Cough Detection (CoughVID)
FPGA Platform	Artix-7	Artix-7	Artix-7	Artix-7
Input Dimension	88200 x 1	220500 x 1	220500 x 1	66150 x 1
Model Size (Kb)	357	198	320	359
Computations (GOP)	2.4	0.6	6	1.8
Fixed Point Precision	8-bit	8-bit	8-bit	8-bit
#PE used	8	8	8	8
Frequency (MHz)	80	80	80	80
Latency (s)	2.3	0.4	3.41	2
BRAM (Used %)	81 (60%)	81 (60%)	81 (60%)	81 (60%)
Total Power (mW)	244	240	245	244
Energy (mJ)	561	96	836	488
Performance (GOPS)	1	1.5	1.8	0.9
Efficiency (GOPS/W)	4.1	6.3	7.3	3.7

- Xilinx Vivado 2018.3 tool for synthesis and implementation.
- The Kintex-7 160t has 107 mW of static power dissipation which is suitable for low power embedded device implementation.

Conclusions and Future Work

UMBC

- The final goal is to develop an integrated AI-based software framework that serve early detection of Flu-like symptoms with detecting COVID-19 with high reliability and inspire innovative solutions for other medical problems.

- We will use our learned knowledge from both demographic info, feature engineering data, SPO2 sensors' data and Speech Processing task data to feed as input of the AI model.

- We will use existing and/or newly formulated AI models to determine the analytical performance metrics of sensitivity, specificity, reliability and accuracy.