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

# Device Invariant Deep Neural Networks for Pulmonary Audio Event Detection Across Mobile and Wearable Devices

Mohsin Y Ahmed, Li Zhu, Md Mahbubur Rahman, Tousif Ahmed, Jilong Kuang and Alex Gao

*Abstract*— **Mobile and wearable devices are being increasingly used for developing audio based machine learning models to infer pulmonary health, exacerbation and activity. A major challenge to widespread usage and deployment of such pulmonary health monitoring audio models is to maintain accuracy and robustness across a variety of commodity devices, due to the effect of device heterogeneity. Because of this phenomenon, pulmonary audio models developed with data from one type of device perform poorly when deployed on another type of device. In this work, we propose a framework incorporating feature normalization across individual frequency bins and combining task specific deep neural networks for model invariance across devices for pulmonary event detection. Our empirical and extensive experiments with data from 131 real pulmonary patients and healthy controls show that our framework can recover up to 163.6% of the accuracy lost due to device heterogeneity for four different pulmonary classification tasks across two broad classification scenarios with two common mobile and wearable devices: smartphone and smartwatch.**

*Clinical relevance*— **The methods presented in this paper will enable efficient and easy portability of clinician recommended pulmonary audio event detection and analytic models across various mobile and wearable devices used by a patient.**

## I. Introduction

Obstructive pulmonary diseases like asthma, chronic obstructive pulmonary disease (COPD) and infectious lung diseases like pneumonia and COVID-19 are among the major health concerns in recent times, causing worldwide pandemic, hospitalizations, and death. Only the novel coronavirus induced ongoing pandemic disease COVID-19 have caused more than 34 million infections resulting in more than 610,000 deaths in the US so far [1]. Obstructive airway diseases like asthma and COPD affect up to 15% of adults in the US, causing more than a million hospitalizations [2].

Most obstructive or infectious pulmonary diseases are characterized by coughing, wheezing due to narrowed lung airway, or altered speech pattern. Smartphones and wearables like smartwatches are the two most widely used commodity devices, with 285 million [3] smartphone and 60.6 million [4] wearable users in 2020 in the US. With such proliferation of mobile and wearable devices among the general population in recent years, various attempts have been made to evaluate pulmonary health [5]–[7], pulmonary exacerbation detection [8], and pulmonary audio analysis [9], [10] using these devices.

A major challenge to widespread usage and deployment of pulmonary health monitoring audio models is to maintain accuracy and robustness across a variety of devices. Due to a phenomenon of device heterogeneity [11], a machine learning model developed with data obtained from one device (like a smartphone) is often not effective if deployed in a different device (like a smartwatch). We have demonstrated this problem empirically in Section II. Developing an effective pulmonary health monitoring model requires massive effort and cost for protocol development, IRB approval, patient and healthy control recruitment, pulmonary audio data collection through controlled and longitudinal deployment, data annotation and annotation verification. If developed models are not portable and usable to newer and upcoming devices, all these steps have to be repeated to collect data with every new device, incurring effort, time and monetary overhead. Each new episode of data collection and annotation may cost between $10,500 to $85,000 [12], depending on the nature of data and the complexity of annotations.

To address these challenges, we present a solution to reduce the data distribution domain shift due to device heterogeneity, and thus recover the lost accuracy of pulmonary event detection audio models when deployed in different devices. At the feature level, we propose to utilize different normalization functions at different frequency bins to transform the original features from the two devices to minimize the effect of heterogeneity. In addition, we incorporate pulmonary task specific deep neural networks (DNN) which in addition to learning discriminative features of various pulmonary events, are capable to learn to perform inference in a device agnostic manner, thus further reducing the effect of device heterogeneity.

The contributions of this work are:

- Novel application of normalization functions on device specific pulmonary audio features by frequency bin transformation and adaptation of audio features, and reduce the effect of domain shift for better model portability across wearable and mobile devices.
- Combining the frequency bin normalized features along with pulmonary event detection task specific DNNs, to learn task specific discriminative representations along with inference capability tolerating the effect of device heterogeneity.
- A comprehensive evaluation of our framework using an extensive and empirical pulmonary activity dataset collected from 131 patients and healthy subjects in four different pulmonary classification tasks across two broad classification scenarios with two common mobile and wearable devices: smartphone and smartwatch.

*M. Y. Ahmed, L. Zhu, M. M. Rahman, T. Ahmed, J. Kuang and A. Gao are with the Digital Health Lab at Samsung Research America, Mountain View, CA, USA. Correspondence: mohsin.ahmed@samsung.com

## II. Problem Demonstration

### A. Frequency Response Heterogeneity

Cross device pulmonary event detection by mobile (smartphones) and wearable (smartwatches) devices are subject to device heterogeneity [11], [13], [14]. Before a pulmonary audio signal even reaches the audio classifier in the device, it goes through several components of the audio processing pipeline consisting of the device microphone hardware and digital signal processing (DSP) software. Wearable devices like smartwatches are generally more resource-constrained than mobile devices like smartphones. To meet user requirements and address energy-constraints, a smartwatch may capture and process audio slightly differently than a smartphone at the hardware and software level. [11] showed that signal heterogeneity across various mobile devices due to their hardware are computationally significant enough to fingerprint smart devices based on their audio signatures. In addition, device microphone manufacturers often set certain device specific parameters which direct the DSP software running on top of it. Since the fingerprintings in [11] are done from different models and manufacturers of smartphones only, we argue that in the broader context of smartphones and smartwatches, such heterogeneity are likely to be even more prominent due to the differences in manufacturing, usage scenario and device resources.

From the perspective of pulmonary event detection models, the heterogeneity of audio signatures across mobile and wearable devices introduces domain shift in the captured signal spectrum across devices. Due to domain shift, the training data from the data collection device will have a different distribution than that of the inference data captured in the deployment device, causing poor inference performance at the target device, regardless of good inference performance for homogeneous data from the native device.

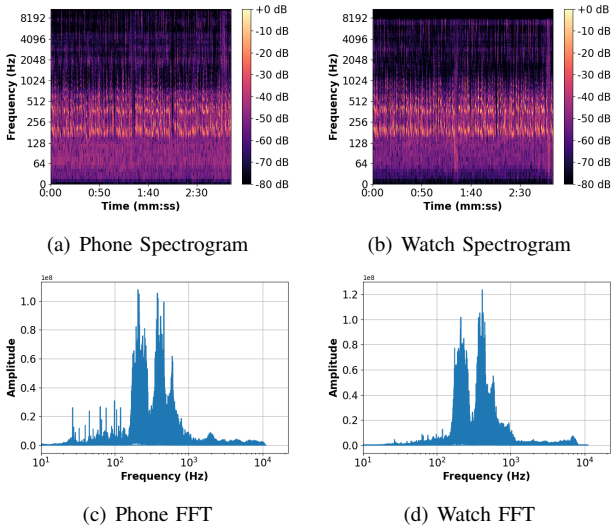
### B. Effect on Frequency Response

In Figure 1, we show spectrograms (short-term Fourier transform with window size = 2048 and hop length = 512) of a speech segment (3 minutes 5 seconds duration) as recorded by a Samsung Galaxy Note 8 smartphone and a Galaxy Active2 smartwatch simultaneously. We observe that the devices exhibit differences in their frequency responses to the exactly same speech input, which are visualized both in the spectrograms and fast Fourier transform (FFT). For example, the data captured from smartwatch has low power in the lower frequency ranges, which infers that either the smartwatch has a physical limitation in its microphone hardware which does not allow it to capture low frequencies well (hardware effect) or they are being filtered out in the software by the microphone digital signal processor (DSP). Next, we discuss how these variabilities in the frequency domain can impact the performance of pulmonary event classifiers across mobile and wearable devices.

### C. Effect on Features and Audio Classifiers

We now demonstrate how device heterogeneity across smartphone and smartwatch can impact pulmonary event detection features and classification models across devices. We record a scripted speech by several people using a Samsung smartphone and smartwatch. The recordings are done simultaneously on both devices in a controlled environment. We extract 20 mel-frequency cepstral coefficient (MFCC) features from both datasets, and notice that, variations in the frequency spectrum are reflected in the audio features too. As seen in Figure 2 box-plot, the distribution of MFCC-1 features from the smartphone is quite different than that of the smartwatch. We train a speaker identification model with data from one device, and perform an inference experiment by testing on data from the other device, and vice versa. As expected, domain shift due to device heterogeneity caused poor performance of the speaker identification classifiers built on top of these differently distributed features. As such, this represents a scenario where there is domain shift between the training and test data distributions.

Table I shows the inference accuracy on each device-specific test set. We observe that although the model has a higher accuracy on the native device dataset, when deployed on the other device, the accuracy drops sharply (62% when deployed on watch and 79% when deployed on phone). This demonstrates that the data distribution mismatch between training and test devices has a severe impact on pulmonary audio models. Therefore, it is important to tackle this challenge to build robust and usable device invariant pulmonary event detection models across mobile and wearable devices.




(a) Phone Spectrogram (b) Watch Spectrogram

(c) Phone FFT (d) Watch FFT

Fig. 1. Spectrogram (1(a) and 1(b)) and FFT (1(c) and 1(d)) of a speech segment recorded in a smartphone (left) and smartwatch (right). Device heterogeneity causes the smartwatch signal having low power in the lower frequency ranges, which results in slightly different frequency signatures of the same signal across the two devices.

TABLE I
Accuracy of cross-device speaker identification across smartphone and smartwatch

| Training Device | Phone Test Accuracy | Watch Test Accuracy |
|---|---|---|
| Phone | 81% | 19% |
| Watch | 19% | 97% |

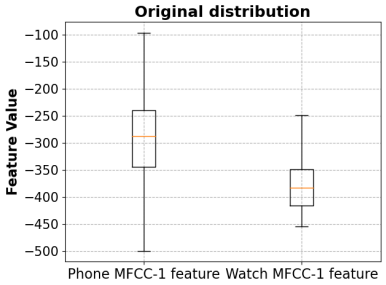



Fig. 2. Box-plot of a feature distribution of a reading segment simultaneously recorded in a smartphone (left) and a smartwatch (right).

## III. Pulmonary Activity Data Collection

We collected real-world pulmonary activity audio data from 131 pulmonary patients and healthy controls using a Samsung Galaxy Note 8 smartphone and a Samsung Active2 smartwatch, at 44.1 KHz sampling rate and 32 bits per sample for both devices. The subject population consisted of total 91 chronic pulmonary patients and 40 healthy subjects, with 69 of them being asthma patients, 9 being COPD patients, and 13 exhibiting co-morbidity of both asthma and COPD. The data collection protocol was reviewed and approved by Institutional Review Board (IRB).

The data collection protocol consisted of each subject performing the following tasks while holding the phone by their chest and the watch by their abdomen, while audio data were recorded by both devices: 1) pulmonary function test (3 efforts), 2) cough naturally/voluntarily for 2 minutes, 3) making A-vowel ('Aaaa...') sound for as long as possible, 4) speaking freely on any topic for 3-5 minutes, 5) reading a neutral paragraph for 3-5 minutes. These tasks were chosen carefully by expert researchers and medical professionals to make the dataset with high quality and enough variability in terms of disease specific pulmonary events. On average, there was around 40 minutes of continuous audio data collected per subject.

## IV. Methods

We define *alien device* as the device which is used for data collection for pulmonary event detection, and we define *host device* as the device where the pulmonary event detection model will be deployed.

### A. Frequency Bin Normalization

The purpose of normalization is to reduce the mismatch in frequency signature across devices due to domain shift. This consists of two core steps: i) alien feature transformation, and ii) host feature adaption.

At *alien feature extraction* stage, $N$ frequency domain features, $f_i(alien), (i = 1...N)$ corresponding to $N$ frequency bins (for example, 20 MFCC features from 20 continuous bins in the frequency spectrum, where $N = 20$) are extracted from each sample data of the alien device.

At *alien feature transformation* stage, a set of statistics for each frequency bin, $S_{f_i(alien)}$ across all training samples from alien device are calculated. Each frequency bin feature of each sample of alien device is transformed using a function **H** (described in Section IV-B) of $f_i(alien)$ and $S_{f_i(alien)}$,

$$f'_i(alien) = \mathbf{H}[f_i(alien), S_{f_i(alien)}], (i = 1...N) \quad (1)$$

At *training* stage, the transformed features from alien device are used to train the appropriate classification model for the pulmonary event detection application.

At *inference* stage, audio to classify is captured using the host device during application run time. The same N frequency domain features $f_i(host), (i = 1...N)$ are extracted from all samples. At *host feature adaptation* stage, each frequency bin feature of each sample of host device is adapted using function **H** of $f_i(host)$ and previously computed statistics $S_{f_i(alien)}$,

$$f'_i(host) = \mathbf{H}[f_i(host), S_{f_i(alien)}], (i = 1...N) \quad (2)$$

Since (1) and (2) both use $S_{f_i(alien)}$ from the training data of the alien device, the frequency distribution at every bin becomes normalized across the two devices, thus minimizing the effect of domain shift. Finally, the adapted host features are classified based on classification model created from transformed features from the alien device.

### B. Normalization Function *H*

We use the following three normalization algorithms:

*1) Z-normalization:* $S = \{\mu_i(alien), \sigma_i(alien)\}$ i.e. mean and standard deviation of each frequency bin of alien features are used along with the Z-transformation as **H** as following for alien feature transformation and host feature adaptation:

$$f'_i(alien) = \frac{f_i(alien) - \mu_i(alien)}{\sigma_i(alien)}, (i = 1...N) \quad (3)$$

$$f'_i(host) = \frac{f_i(host) - \mu_i(alien)}{\sigma_i(alien)}, (i = 1...N) \quad (4)$$

*2) Min-Max Normalization:* $S = \{min_i(alien), max_i(alien)\}$ i.e. minimum and maximum of each frequency bin of alien features are used as following **H** for alien feature transformation and host feature adaptation:

$$f'_i(alien) = \frac{f_i(alien) - min_i(alien)}{max_i(alien) - min_i(alien)}, (i = 1...N) \quad (5)$$

$$f'_i(host) = \frac{f_i(host) - min_i(alien)}{max_i(alien) - min_i(alien)}, (i = 1...N) \quad (6)$$

*3) Quartile/Robust Normalization:* First, second, and third quartile of each frequency bin across all training samples from the alien device, i.e. $S = \{Q1_i(alien), Q2_i(alien), Q3_i(alien)\}$, can be used as following **H** for alien feature transformation and host feature adaptation:

$$f'_i(alien) = \frac{f_i(alien) - Q2_i(alien)}{Q3_i(alien) - Q1_i(alien)}, (i = 1...N) \quad (7)$$

$$f'_i(host) = \frac{f_i(host) - Q2_i(alien)}{Q3_i(alien) - Q1_i(alien)}, (i = 1...N) \quad (8)$$

### C. Deep Neural Networks for Device Invariance

Once we reduce the domain shift at the feature level across the devices, the final component of our pulmonary event detection framework is a deep neural network. The DNN has two objectives: i) to learn discriminative representations based on the pulmonary classification task at hand (e.g., recognizing whether a patient is coughing) and (2) perform inference of pulmonary events further minimizing the effect of domain shift due to software and hardware heterogeneity across smartphone and smartwatch. Our major novelty is bridging the frequency bin normalized features along with the DNN, which we call as *normalized DNN*, and to achieve the above mentioned dual objectives. The final outcome of such *normalized DNN* is a heterogeneity-tolerant device-invariant deep model for pulmonary event detection that can be used across mobile and wearable devices.

## V. Evaluation

### A. Methodology

*1) Pulmonary Classification Scenarios:* Our experiments are done for the following two scenarios of pulmonary event detection tasks:

*a) Person Identification from Pulmonary Test Sounds:* The goal here is to perform person identification from pulmonary test audios. Audio analysis of cough, speech and pulmonary function tests in mobile and wearable devices are shown to be effective for pulmonary health assessment and severity estimation of pulmonary diseases [8], [15]–[17]. In an uncontrolled free-living recording environment, it is important to ensure that the captured sound in the device indeed belongs to the primary subject, as opposed to unintentionally being recorded by some surrounding subjects in the vicinity. Person identification from pulmonary test sounds ensures collection of pulmonary audio from the primary target subject, and thus avoids erroneous health analysis or disease diagnosis by clinicians or predictive algorithms. We evaluate the efficacy of our DNN model invariance across smartphone and smartwatch for two types of person identification tasks: i) person identification from cough (coughID), and ii) person identifcation from speech (speakerID).

*b) Pulmonary Event Detection:* Pulmonary exacerbation due to narrowed lung airway are often characterized by cough episodes, irregular speech patterns and wheezing. Therefore detection and analysis of these pulmonary events provide valuable insights for pulmonary health analysis, degree of exacerbation and disease severity. We experiment with two types of pulmonary event detection tasks across smartphone and smartwatch: i) cough detection, and ii) speech detection.

*2) DNN Architectures:* Our person identification model is a seven-layer DNN trained by MFCC features extracted using the librosa Python library. The numbers of neurons within the hidden layers are 4096, 2048, 1024, 512, 128, and 64. Six subjects were randomly selected and their scripted speech and coughing sessions recorded simultaneously in the smartphone and smartwatch were used for the person identification experiments. Our pulmonary event detection model is a four-layer DNN trained by MFCC features and consists of three hidden layers with 128, 64 and 64 neurons each. Smartphone and Smartwatch recordings of fifty randomly selected subjects were used for these experiments. Each experiment was repeated several times with new random re-selection of subjects every time for statistical significance. We initialize all network weights using Glorot uniform initialization [18] and categorical cross-entropy [19] was used as the loss function. For regularization, we use L2-regularization with $\lambda = 0.001$, and dropout [20]. Both models were implemented in Keras with Tensorflow backend. The DNN configurations are summarized in Table II.

TABLE II
DNN configuration for the 2 types of tasks: Person Identification from Pulmonary Sounds and Pulmonary Event Detection.

| DNN Configuration | Person Identification (CoughID and SpeakerID) | Pulmonary Event Detection (Cough and Speech) |
|---|---|---|
| *No. of hidden layers* | 6 | 3 |
| *No. of neurons in each layer* | (4096, 2048, 1024, 512, 128, 64) | (128, 64, 64) |
| *Optimizer* | adamx | adam |
| *Activation function* | relu (1st layer) and tanh | relu |
| *Kernel initializer* | glorot uniform | glorot uniform |
| *Loss* | categorical crossentropy | categorical crossentropy |
| *Regularization* | Dropout after 1st layer with rate 0.1 | Dropout after 1st layer with rate 0.2 |

*3) Classifier Baselines:* We compare our normalized DNN framework against two shallow baseline classifiers: Random Forest, and Logistic Regression. To train the shallow baseline classifiers, we used same frequency bin normalized features for each task to act as task-specific classifier baselines.

### B. Performance Gains

*1) Experimental setup:* We present a series of experiments using differing devices to train and test the DNN models using different normalization functions. We compare the performance of normalized DNNs against a *best case* where original data from same device are used for training and testing, and a *baseline case* where original data from one device is used for training while original data from other device is used for testing without any data transformation.

*2) Results:*

*a) Person Identification:* In Figure 3, we present the weighted F1 scores on test devices when the person identification model is trained for best case (same device, original data), baseline case (different device, original data), and normalized DNN (different device, normalized data) on the speech and cough recordings collected from smartphone and smartwatch. For both coughID (3(a), 3(b)) and speakerID (3(c), 3(d)), we observe a sharp degradation of performance in the baseline cases compared to the best cases because of domain mismatch due to device heterogeneity. The accuracy loss is 57.7% and 79.3% for coughID, and 76.5% and 81.4% for speakerID, for phone-train-watch-test and watch-train-phone-test baseline scenarios respectively, compared to corresponding best case scenarios. Incorporating alien data transformation and host data adaptation using

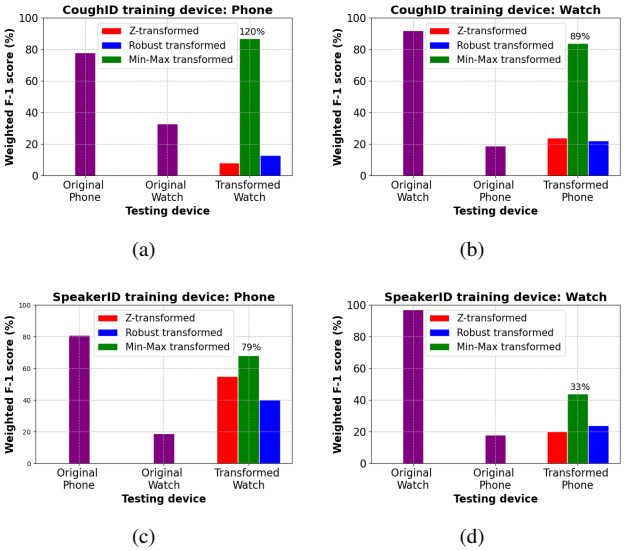



Fig. 3. Effects of different normalization algorithms on coughID (3(a) and 3(b)) and speakerID (3(c) and 3(d)). Min-Max consistently performs better than the other two algorithms in recovering lost accuracy across the devices due to non Gaussian distribution of data. Numbers on top of the green bars denote percentage recovery due to Min-Max normalization.

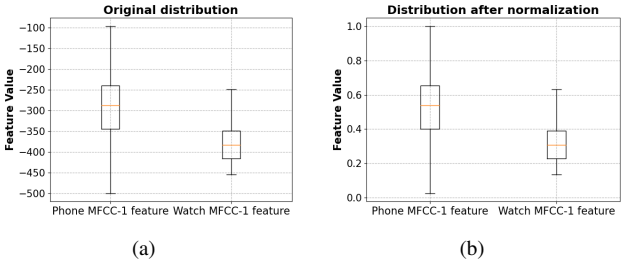



Fig. 4. Min-Max normalization on frequency bins reduces the feature distribution distance between phone and watch.

the normalization techniques significantly recover the lost accuracy due to domain shift, as shown in Figure 3. The Min-Max normalization algorithm significantly outperforms the other two algorithms (Z-normalization and Quartile/Robust normalization), the likely reason being the non-Gaussian nature of distribution of the features. By incorporating Min-Max normalized DNN in the inference pipeline, we were able to recover 120% and 89% of lost accuracy for coughID, and 79% and 33% recovery for speakerID, for phone-train-watch-test and watch-train-phone-test scenarios respectively.

*b) Effect of Normalization:* For demonstrating the effect of normalization on feature distributions across smartphone and smartwatch, we use the speakerID dataset recorded with both devices. Figure 4 shows the comparative box plot distributions of one of the features (MFCC-1) recorded with the two devices before and after the Min-Max normalization is performed on the original features. The post-normalization difference between the medians (50-th percentile) of the feature distributions across the two devices are 0.4, compared to a difference of 173.66 pre-normalization. Similar effect is observed for other MFCC features as well. This demonstrates that Min-Max normalization is able to

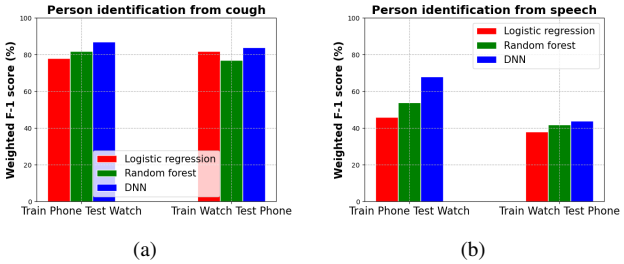



Fig. 5. Comparison of optimized DNN and shallow classifiers with Min-Max normalized features for (a) coughID and (b) speakerID. The DNN was able to learn more device invariant representations than the shallow classifiers beyond the normalized features.

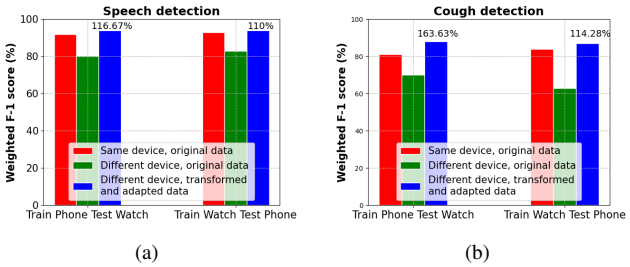



Fig. 6. Performance of (a) speech detection and (b) cough detection across phone and watch with DNN with Min-Max normalized features.

reduce the domain shift between smartphone and smartwatch audio.

*c) Comparison with Shallow Classifiers:* To demonstrate the utility of our DNN classifier compared to some shallow baseline classifiers, we use the coughID and speakerID datasets recorded with both devices with Min-Max normalization, as this was found to be the best normalization algorithm. We use two popular shallow classifiers as comparison baseline: Random Forests [21], and Logistic Regression [22]. Figure 5 shows that, although using the same Min-Max normalized features across devices, our normalized DNN outperforms both the shallow classifiers with the same normalized features for coughID and speakerID tasks as input. This demonstrates that, our normalized DNN was able to learn more device invariant representations for the pulmonary tasks than the shallow classifiers beyond the normalized feature level.

*d) Pulmonary Event Detection:* Finally, Figure 6 shows the results of pulmonary event detection tasks, i.e. cough detection and speech detection using normalized DNN (different device, normalized data), compared with a best case (same device, original data), and baseline case (different device, original data). As expected, there was 13.04% and 10.75% performance drop for speech detection, and 13.58% and 25% drop for cough detection for baseline cases compared to best cases, for phone-train-watch-test and watch-train-phone-test scenarios respectively. Min-Max normalized DNN was able to recover 116.67% and 110% of lost performance for speech detection, and 163.6% and 114.28% of lost performance for cough detection for each of the two cross-device scenarios.

### C. Comparison to Mic2Mic

Mic2Mic [14] is a state-of-the-art system to address device heterogeneity for model robustness across devices. It uses cycle-consistent generative adversarial networks (CycleGAN) [23] to learn pairwise mapping from a source device to target device. We compare the features and benefits of our normalized DNN compared to Mic2Mic in the following:

*1) Translation Cardinality:* Mic2Mic can only learn pairwise and directional translation function across devices using CycleGAN and cycle-consistency property. Therefore, a learned translation function between two devices cannot be used in reverse direction or for other devices, demonstrating a one-to-one unidirectional translation mapping across devices. In contrast, our method is flexible to translate and adapt any alien device data to a given host device data, i.e. demonstrating an any-to-one translation mapping property.

*2) Training Requirement:* Mic2Mic requires an additional step of device-to-device data translation in the inference pipeline of audio models, using CycleGANs which are time consuming and effort heavy to train. On the contrary, our translation function between smartphone and smartwatch requires frequency bin normalization of alien device features and adaptation of host device features, along with the DNN training, which are easier and less time consuming to train compared to Mic2Mic.

*3) Data Requirement:* Mic2Mic requires pre-collected data with every device to train its mapping function with CycleGAN. This is in addition to the data requirement of the end machine learning application, which will require task specific data as well. This adds an additional overhead of data recording in a deployment site. Our approach does not need any extra data collection and only requires the machine learning inference data from the alien device which it transforms using frequency bin normalization and uses the transformation parameters to adapt host device data.

*4) Performance Recovery:* Mic2Mic approach has been able to recover a maximum 87% of lost performance due to domain shift, as reported in [14]. As shown in Section V-B, our normalized DNN have more than 100% recovery in multiple pulmonary activity detection tasks, which means that the feature normalization and DNN training learns cross device mappings so well that it often outperforms the best case with original unaltered data, which Mic2Mic has not been able to do.

## VI. Discussion

### A. Extending to Other Devices

Apart from smartphones and smartwatches, other wearable devices like smart earbuds [24] and IoT devices like virtual assistants (Alexa [25], Siri [26]) have great potential for pulmonary event monitoring. As discussed in Section V-C.1, our normalized DNN approach has an any-to-one translation property from any alien device to the target host device. This property would enable easy portability of already developed pulmonary event detection models from smartphone and smartwatch data to new and upcoming mobile, wearable, and IoT devices.

### B. Phone-Watch Synchronization

Although pulmonary audio data collection was done simultaneously with the smartphone and smartwatch, the final recordings across the two devices were not properly time aligned and had little drift (in orders of few seconds). This drift is due to the unique characteristics of each device's internal clock, unpredictable thread scheduling and data buffer management [27]. Since each of our cough and speech recordings sessions were several minutes long, this time drift of few seconds across the devices were negligible. Therefore, we could use the annotation from only the smartphone to label data from both devices. However, in cases where the original events are too short such that these cross device time drifts become significant, annotation has to be done individually for each device.

### C. Effect on Other Features

We demonstrated the efficacy of normalized DNN for pulmonary event detection using 20 MFCC features only. However, our approach is extendable to address domain shift of any frequency domain acoustic feature. Popular feature extractor tools like OpenSmile [28] is able to extract many other frequency based acoustic features like perceptive linear prediction (PLP) coefficients, perceptive linear prediction cepstral coefficients (PLPCC), line spectral frequencies, fundamental frequency, frequency components, spectral centroid, spectral features (e.g., flux, density, roll-off, arbitrary band energies, centroid, entropy, variance, skewness, kurtosis, slope), formant frequencies, and bandwidths, which have been used widely for a wide class of audio event detection. The phenomenon of domain shift across devices is noticeable in the frequency domain only, therefore this will not affect time domain acoustic features, like energy of signal, zero crossing rate, maximum amplitude, minimum energy etc across devices.

## VII. Related Work

Recent works explored both mobile and wearable audio sensing to detect and analyze pulmonary events. Smartphone has been used for detecting cough [16], [17], [21], wheeze [8]–[10], characterizing the events such as whether a cough is wet or dry [5], and detecting snoring and distinguishing it from nocturnal coughs [29]. Smartphone audio is further used to extract important pulmonary digital biomarkers and assess pulmonary patient conditions [8], [9], [15]. Previous works also showed the feasibility of utilizing detected cough [15]–[17] and speech [15] to estimate lung function or detect chronic pulmonary patients. Smartphones placed near to mouth or nose can be used to estimate breathing and airflow sound for lung obstruction estimation [30], [31].

Very recently, smartwatch based audio sensing showed great potential to passively detect and monitor sound events (e.g., cough) for chronic pulmonary patients. Liaqat et al. [32] showed that cough can be detected using smartwatch in privacy preserving ways by obfuscating speech in the recordings. In a 3-month long study, Wu et al., [33] showed

that some pulmonary patients will wear and maintain smartwatches to passively monitor audio, heart rate, and physical activity, and smartwatches are able to reliably capture near-continuous patient data [34]. All these useful techniques and models developed for pulmonary healthcare are developed and tested using one type of device only, and are not portable to other devices due to domain shift from device heterogeneity.

## VIII. Conclusion

This paper presents a framework for portability of pulmonary activity detection and analysis audio models across mobile and wearable devices using frequency bin normalization and adaptation of device level audio features combined with task specific DNNs. Compared to state-of-the-art methods, our method is training free at feature translation level, does not require any additional data collection for translation between devices, and can translate and adapt any alien device model to a given host device. Empirical experiments using data collected from 131 pulmonary patients and healthy controls using smartphone and smartwatch demonstrates upto 163.6% recovery of lost baseline performance when models are ported across devices. With mobile and wearable devices having great potential for home healthcare and remote patient monitoring, our framework will enable seamless portability of clinician recommended pulmonary models across various consumer devices used by the patient population.

## References


[1] Covid-19 statistics. https://covid.cdc.gov/covid-data-tracker.

[2] Stavros Garantziotis and David A Schwartz. Ecogenomics of respiratory diseases of public health significance. *Annual review of public health*, 31:37–51, 2010.

[3] Smartphone usage. https://www.statista.com/statistics/201182/forecast-of-smartphone-users-in-the-us.

[4] Wearable usage. https://www.statista.com/statistics/543070/number-of-wearable-users-in-the-us.

[5] Ebrahim Nemati et al. A comprehensive approach for classification of the cough type. In *EMBC*, pages 208–212. IEEE, 2020.

[6] Md Mahbubur Rahman et al. Breatheasy: Assessing respiratory diseases using mobile multimodal sensors. In *ICMI*, pages 41–49, 2020.

[7] Viswam Nathan et al. Extraction of voice parameters from continuous running speech for pulmonary disease monitoring. In *BIBM*, pages 859–864. IEEE, 2019.

[8] Soujanya Chatterjee et al. Assessing severity of pulmonary obstruction from respiration phase-based wheeze-sensing using mobile sensors. In *CHI*, pages 1–13, 2020.

[9] Mohsin Y Ahmed et al. mlung: Privacy-preserving naturally windowed lung activity detection for pulmonary patients. In *BSN*, pages 1–4. IEEE, 2019.

[10] Mohsin Y Ahmed et al. Deeplung: Smartphone convolutional neural network-based inference of lung anomalies for pulmonary patients. In *INTERSPEECH*, pages 2335–2339, 2019.

[11] Anupam Das, Nikita Borisov, and Matthew Caesar. Fingerprinting smart devices through embedded acoustic components. *arXiv preprint arXiv:1403.3366*, 2014.

[12] Cost of data collection. https://medium.com/cognifeed/the-cost-of-machine-learning-projects-7ca3aea03a5c.

[13] Akhil Mathur et al. Using deep data augmentation training to address software and hardware heterogeneities in wearable and smartphone sensing devices. In *IPSN*, pages 200–211. IEEE, 2018.

[14] Akhil Mathur et al. Mic2mic: using cycle-consistent generative adversarial networks to overcome microphone variability in speech systems. In *IPSN*, pages 169–180, 2019.

[15] Keum San Chun et al. Towards passive assessment of pulmonary function from natural speech recorded using a mobile phone. In *PerCom*, pages 1–10. IEEE, 2020.

[16] Ebrahim Nemati et al. Estimation of the lung function using acoustic features of the voluntary cough. In *EMBC*, New York, NY, USA, 07 2020. IEEE, IEEE.

[17] Vishwajith Ramesh et al. Coughgan: Generating synthetic coughs that improve respiratory disease classification. In *EMBC*, New York, NY, USA, 07 2020. IEEE, IEEE.

[18] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256, 2010.

[19] Zhilu Zhang and Mert R Sabuncu. Generalized cross entropy loss for training deep neural networks with noisy labels. *arXiv preprint arXiv:1805.07836*, 2018.

[20] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

[21] Ebrahim Nemati et al. Private audio-based cough sensing for in-home pulmonary assessment using mobile devices. In *EAI International Conference on Body Area Networks*, pages 221–232. Springer, 2018.

[22] Mohsin Y Ahmed et al. Socialsense: A collaborative mobile platform for speaker and mood identification.

[23] Ian J Goodfellow et al. Generative adversarial networks. *arXiv preprint arXiv:1406.2661*, 2014.

[24] Chulhong Min, Akhil Mathur, and Fahim Kawsar. Exploring audio and kinetic sensing on earable devices. In *Proceedings of the 4th ACM Workshop on Wearable Systems and Applications*, pages 5–10, 2018.

[25] Alexa. https://developer.amazon.com/en-US/alexa.

[26] Siri. https://www.apple.com/siri.

[27] Tousif Ahmed at al. Automated time synchronization of cough events from multimodal sensors in mobile devices. In *ICMI*, pages 614–619, 2020.

[28] Florian Eyben, Martin Wöllmer, and Björn Schuller. Opensmile: the munich versatile and fast open-source audio feature extractor. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 1459–1462, 2010.

[29] Sudip Vhaduri et al, Theodore Van Kessel, Bongjun Ko, David Wood, Shiqiang Wang, and Thomas Brunschwiler. Nocturnal cough and snore detection in noisy environments using smartphone-microphones. In *ICHI*, pages 1–7. IEEE, 2019.

[30] Md Mahbubur Rahman et al. Exhalesense: Detecting high fidelity forced exhalations to estimate lung obstruction on smartphones. In *PerCom*. IEEE, 2020.

[31] Mayank Goel et al. Spirocall: Measuring lung function over a phone call. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 5675–5685, 2016.

[32] Daniyal Liaqat et al. A method for preserving privacy during audio recordings by filtering speech. In *2017 IEEE Life Sciences Conference (LSC)*, pages 79–82. IEEE, 2017.

[33] D Liaqat et al. Towards ambulatory cough monitoring using smartwatches. In *C41. HEALTH SERVICES RESEARCH IN PULMONARY DISEASE*, pages A4929–A4929. American Thoracic Society, 2018.

[34] Robert Wu et al. Feasibility of using a smartwatch to intensively monitor patients with chronic obstructive pulmonary disease: prospective cohort study. *JMIR mHealth and uHealth*, 6(6):e10046, 2018.