A Scalable and Domain Adaptive Respiratory Symptoms Detection Framework using Earables

Nhan Nguyen
Department of Electrical and Computer Engineering
Texas A&M University
College Station, USA
Email: nhannguyen993@tamu.edu

Avijoy Chakma, Nirmalya Roy
Department of Information Systems
University of Maryland, Baltimore County
Baltimore, USA
Email: {achakma1, nroy}@umbc.edu

2021 IEEE International Conference on Big Data (Big Data) | 978-1-6654-3902-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/BigData52589.2021.9671796

***Abstract*—The COVID-19 pandemic has brought a devastating impact on human health across the globe, and people are still observing face-masking as a preventive measure to contain the spread of COVID-19. Coughing is one of the major transmission mediums of COVID-19, and early cough detection could play a significant r ole i n p reventing t he s pread o f t his life-threatening virus. Many approaches have been proposed for developing systems to detect coughing and other respiratory symptoms in literature, but earable devices are not well-studied and investigated for respiratory symptom detection. In this work, we posited an acoustic research prototype (earable device) - *eSense* that has acoustic and IMU sensors embedded into user-convenient earbuds to address the following issues: (i) feasibility of the earables in detecting respiratory symptoms, and (ii) scalability of trained machine learning models in the presence of unseen data samples. We performed experimentation with both shallow and deep learning models on the *eSense* collected data samples. We observed that the deep learning model outperforms the shallow learning models achieving 97% accuracy. Furthermore, we investigated the scalability of the deep learning model on unseen datasets and noticed that the performance of the deep learning model deteriorates when trained on a particular dataset and tested on an unseen dataset. To mitigate such challenges, we postulated an adversarial domain adaptation technique that helps improve the performance of our respiratory symptoms detection framework by a substantial margin.**

***Index Terms*—Domain Adaptation, eSense, COVID-19, Respiratory Symptoms, Earables, Smart Health**

I. Introduction

Internet of Things (IoT) advancement has created a wide-range innovation of smart devices enhancing human capability in developing tools towards smart cities, smart homes, energy expenditure monitoring, fitness t racking, a nd s mart health applications. Earable devices are the recent addition of IoT technologies that enhances earbud capability to interact and respond with the surrounding environment through various sensor functionalities and empower the end-user with various applications. Such smart earbuds hold the benefits o f portability, user-friendliness, and unobtrusiveness which helps to obviate unrealistic assumptions of device-use scenario. *eSense* is such an earable device, which is composed of acoustic, IMU, and BLE sensors [1]. Several literature works have studied its feasibility on smart health applications [2], [3] but till now it has not been studied for detecting respiratory symptoms, such as coughing which plays a huge role in spreading the COVID-19 virus.

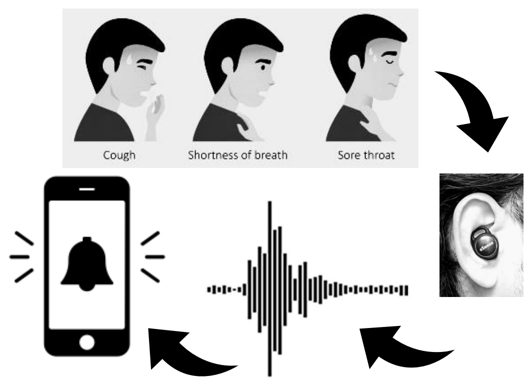

Fig. 1. Respiratory symptoms data collection by the *eSense* earable with the captured signals transmitted to a Bluetooth-connected smartphone. The accumulated data is processed offline to detect the respiratory symptoms and evaluate domain adaptation feasibility.

The devastating effect of COVID-19 across the globe has motivated us to investigate the application of such devices in detecting respiratory symptoms such as coughing, sore throat, and sneezing, as these symptoms are identified as COVID-19 symptoms by the World Health Organization. However, even though in literature, detection of these symptoms is well-explored [4]–[6], the majority of these approaches pose inherent limitations in terms of user-convenience, portability, device-placement-induced noise in data collection, and operation environment. In this context, the *eSense* earable prototype obviates these problems, but the feasibility of such devices in detecting the mentioned symptoms is unknown. As the feasibility with earables has not been well-studied there is no pre-existing dataset that can be easily experimented with. On the other hand, the data collection process poses two major challenges: 1) due to the sensitivity of the COVID-19 pandemic, it is impractical to manually collect voluntary coughing data from a large number of participants, and 2) the crowd-sourcing-based voluntary data collection process incurs logistic challenges over the required procedures and the setup.

In addition to the feasibility study, we also attempted to improve the model performance when the model is exposed to the unseen data sample through the application of domain adaptation techniques - a transfer learning technique primarily used to find out the labels of an unlabeled dataset with the help of an existing labeled dataset. In general, the data distributions of different datasets are different even if the intended task is similar. Such data distribution heterogeneity occurs mainly due to the on-body device placement variations, the difference in the device sampling frequency, end-user sensitivity in the performed activities, and activity execution habits and patterns. More about the domain adaptation is covered in section II. Overall, in an effort to find out the answers on the earable device feasibility on respiratory symptom detection and to appease the model scalability issue, the followings are the primary contributions of this proposed work-

1) **Respiratory Symptom Detection Feasibility Study Using *eSense*** We collected a voluntary respiratory symptom dataset from 4 participants for our feasibility study as depicted in Figure 1 and evaluated the symptom detection performance under various machine learning models (under the best performance settings). We collected voluntary coughing, sore throat, and sneezing data in the collected dataset. Our study reveals the deep learning model efficiency over the traditional machine learning models for earable data processing.
2) **Model Scalability Through Domain Adaptation Technique to Mitigate Performance Degradation** The lack of publicly available datasets leads us to explore the application of domain adaptation techniques on earable data. In general, the data distribution heterogeneity of different datasets causes performance degradation even if the intended dataset use cases are similar. We investigated such performance degradation and applied a domain adaptation technique to mitigate the degradation. We crawled a publicly available dataset and utilized it in our domain adaptation study.

The paper is organized as follows. In section II we discuss the recent works on the *eSense* earable, acoustic cough detection, and domain adaptation techniques. Section III presents our proposed domain adaptation deep learning architecture. We articulate our collected datasets, data preprocessing, ablation study, and the experiments in section III. We discuss the experiments, findings, and implementation aspects in sections IV, V, and VI respectively. Finally, we conclude our study in section VII.

II. Related Work

Respiratory Symptom Detection Several systems have been proposed and developed for respiratory symptom detection from acoustic signals [4], [7]. [7] proposes a real-time low-power wireless respiratory monitoring system to measure the breathing rate and coughing frequency. [4] have developed a wireless sensing system capable of detecting voluntary coughs, sneezes, and face touching that uses radio frequency technology. A deep learning-based approach is proposed by [5] that employs wearable acoustic sensors for cough detection. [8] have evaluated various hand-crafted acoustic features such as SIFT, MFCC, MFB for cough detection using deep architectures and found that treating the cough signals as a single input feature instead of multiple shorter features provides better performance. These approaches offer innovative solutions and insights but incur several drawbacks in terms of the system requirement [4], [7], user-friendliness [7], on-body placement requirement [7], and practicality issues [9] where many of such limitations are obviated by an earable smart device like *eSense* [1] that does not require an expensive setup which resolves the mentioned issues.

***eSense* Literature Work** Recently, many health-related applications have been explored using *eSense* earables. [10] have examined a case study of acoustical manipulation in a blindfold walking scenario on both subtle and overt conditions. [2] finds the effectiveness of the embedded inertial measurement units (IMUs) inside earphones that offer a clear advantage in step counting. [3] proposes a system for sensing respiratory rates using in-ear headphone (*eSense*) inertial measurement units (IMU) that obviates specialized equipment for the same purpose. Our study leverages the *eSense* earable and compliments the traditional coughing as well as respiratory symptom detection research findings.

Domain Adaptation Domain adaptation is a transfer learning approach that aims to label the unlabeled data source with the help of an existing labeled data source. It is assumed that the intended task (classification/regression) using the labeled and unlabeled data sources are the same. In general, there are a few approaches to achieving domain adaptation, and learning the generalized feature space of both datasets is one of them. For more details, we refer the readers to the existing literature work on domain adaptation [11], [12] for detailed definitions, methodologies, techniques. RevGrad [13] proposes an adversarial-based domain adaptation mechanism to extract the generalized features between different data sources for a similar classification task. The challenge we have faced at the beginning of our study with the limited labeled data samples prompted us to explore the domain adaptation approach, more specifically RevGrad [13], to improve our model scalability. Along with the earable feasibility study for respiratory symptom detection, we also evaluated the scope for domain adaptation by:

- Training the deep model using our collected dataset and following by testing on the publicly available coughing dataset (*performance degrades*)
- Improving the performance by integrating a similar adversarial mechanism as the one proposed by RevGrad [13]

In the next section, we describe our proposed deep model with the integrated domain adaptation module.

III. Methodology

Our overall methodology is two folds - firstly, we experimented with several traditional machine learning models and a CNN-based deep model to investigate the feasibility of earables in respiratory symptom detection. In the second fold, we take inspiration from RevGrad [13] and adopt a domain adaptation mechanism to increase our deep learning-based model scalability. We describe the traditional machine learning model settings and different experiments in sub-section IV-C.

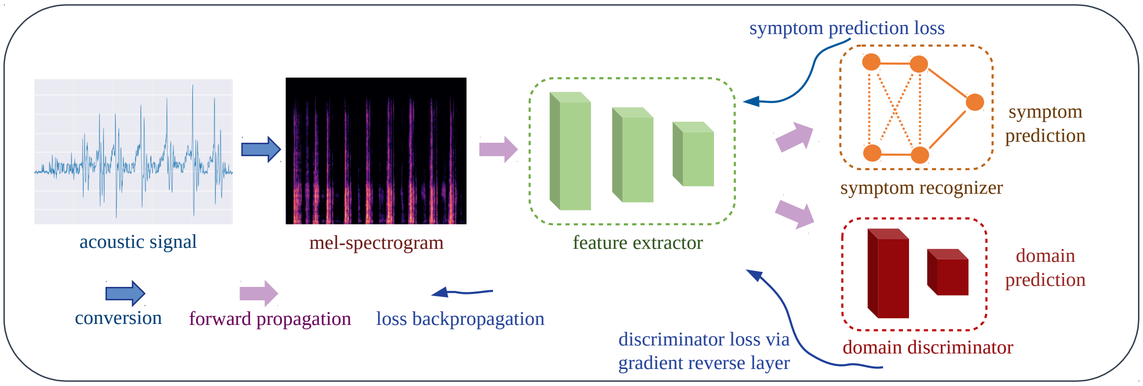

Fig. 2. Overall process of respiratory symptom detection. The acoustic signal is processed into a Mel-spectrogram image. The convolutional neural network is used for feature extraction and a fully connected layer is used for inferencing the symptoms.

In this section, we elaborate on the deep learning network architecture and the adversarial training mechanism.

A. Architecture

The overall deep learning framework is depicted in Figure 2 can be split into three modules - feature extraction (green color), symptom recognizer (orange color), and domain discriminator (maroon color). The feature extraction module consists of 3 layers of convolutional neural network (CNN), Rectified Linear Unit (ReLU), and Max-pooling layers, whereas the inference module consists of 2 fully connected layers. Finally, the domain adaptation module is similar to the inference module but differs in the final layer output count. Details of the architecture settings are tabulated in Table I.

TABLE I
ARCHITECTURE HYPER-PARAMETERS

Hyper-parameters	Values
Convolution Filter No.	64, 128, 256
Convolution Filter Dimension	9x9, 7x7, 3x3
Max Pooling Filter, Stride	2x2, 2
No. of Units in Inference Module	128, 3
No. of Units in Domain Adaptation Module	128, 1

B. Training Mechanism

Initially, we have used the feature extractor and symptom recognizer for the initial feasibility study. Using our labeled collected dataset, we train the network in a supervised manner using a batchwise categorical cross-entropy loss function as described in Equation 1

$$L_{cce} = -\sum_{i=1}^{C} t_i log(p_i) \quad (1)$$

where C is the total number of classes (respiratory symptoms), t_i is the ground truth and p_i is the softmax probability of the symptom recognizer.

We extend our deep learning model for symptom detection for the domain adaptation purpose to improve the model scalability. The trained model in the previous state are further leveraged and adopted for the adaptation phase. During the domain adaptation phase, the deep learning architecture is extended through integrating a domain discriminator (maroon color in Figure 2). The domain discriminator operates in an adversarial manner where it receives the feature from both the labeled and unlabeled datasets. The domain discriminator attempts to predict the source originality of the incoming features whereas the feature extractor operates to negate this goal. We refer to the labeled dataset as the source dataset and the public dataset as the target dataset where the label is assumed to be unavailable. During the adaptation process, feature extractor and domain discriminator are used for both the labeled and unlabeled data samples whereas the symptom recognizer is used only for the labeled data samples. In this work, we have adopted the gradient reversal layer implementation (Equation 4, 5, 6) from the RevGrad [13] work.

IV. EXPERIMENTS

In our symptom detection feasibility study, we employ 4 traditional machine learning models (random forest, decision tree, support vector machine, and multi-linear perceptrons), and a deep learning model. The deep learning model is further extended with the inclusion of a domain discriminator for the domain adaptation feasibility study. We will describe various experimental steps involved with these models including dataset, pre-processing, implementation details, experiment design, and findings throughout our experiments in detail.

A. Dataset

In-House Dataset: We collect voluntary data from 4 participants for three respiratory symptoms (coughing, sore throat, sneezing) in the home environment. All of the participants are male and graduate students aged between 23-30 years old. We collect the data in segments and instruct the participants to provide 5 seconds of data so as not to cause any health issues. We collect 15 seconds of data for each symptom from each participant. We leverage an android app[1] that facilitates a

[1] https://www.esense.io/

connection with *eSense* and collects respiratory symptom data in the android device storage. The participants are allowed to follow their natural pattern while expressing symptoms and precautionary COVID-19 measures are followed during the whole data collection process. The data distribution of our in-house dataset is shown in Figure 3.

Virufy COVID-19 Open Cough Dataset: The lack of sufficient data in our in-house dataset creates additional challenges and motivates us to investigate and apply domain adaptation techniques. We have found one publicly available crowd-sourced coughing dataset that contains COVID-19 audio samples from 16 patients and the corresponding PCR test status. As a crowd-sourced dataset, the dataset is diverse in terms of country of origin, gender, age, and medical history which is perfectly suitable for the domain adaptation study. More details about the dataset can be found in the GitHub repository [2]

B. Dataset Preprocessing and Implementation Details

We use *librosa* [14] and *python_speech_features* [15] library for acoustic signal preprocessing. The deep-learning architecture is implemented using the open-source *PyTorch* [16] library. Statistical features similar to [17] are extracted for the traditional machine learning models using the python-based library - *tsfresh* [18]. In our experiments, we have used the python package *scikit-learn* [19] to implement the machine learning algorithms.

Acoustic data processing steps involve segmenting the acoustic data and generating Mel-spectrograms as the input for the deep model. We split the audio signals into non-overlapping segments of various lengths (0.5 sec, 1 sec). To avoid an extra data preprocessing step, silent sections are not removed from the acoustic files. We execute the experiments on a Linux Server (Ubuntu 18.04) running on Intel Core I7-6850K CPU and 64GB DDR4 RAM, with 4 Nvidia 1080Ti Graphics cards containing 44GB VRAM. We report the accuracy score on the symptom classification task.

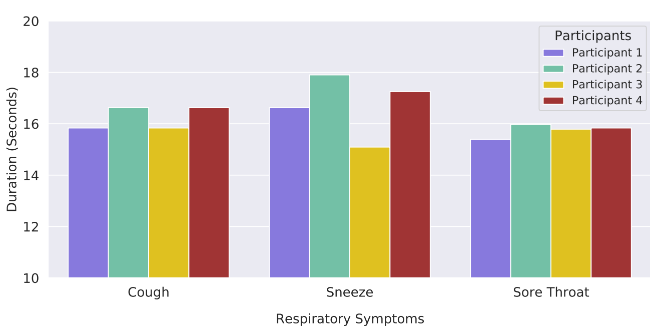

Fig. 3. In-House Data Distribution

C. Model Hyper-parameter Tuning

We fine-tune the model parameters through exhaustive experiments. In the traditional machine learning models, we follow a Leave-One-Subject-Out (LOSO) evaluation where a model is trained on three different users' data and tested on the fourth user's data. We repeat this procedure for each model under different parameter values to find out the best-performing settings as reported in Table II.

Optimum settings for the deep model can be split into two phases. Initially, we experiment with different audio segment lengths and input Mel-spectrogram image sizes for the deep learning model. This experiment is followed by experiments with the network hyper-parameters. Experimental findings are reported in Table III and Table IV respectively.

TABLE II
OPTIMAL PARAMETERS FOR MACHINE LEARNING MODELS

Model	Parameters Values
Random Forest	$n\_estimators = 20$, $max\_depth = 2$, $min\_samples\_split = 0.3$
SVM	$c = 100$, $kernel = rbf$, $gamma = scale$
MLP	$activation = tanh$, $hidden\_layer\_sizes = 100$
Decision Tree	$criterion = entropy$, $max\_depth = 10$

TABLE III
PERFORMANCE UNDER VARIOUS AUDIO SEGMENT LENGTHS AND MEL SPECTROGRAM IMAGE SIZES

Audio Segment	Image Size (Width x Height)		
	$(224x224)$	$(512x512)$	$(1024x1024)$
0.5 Sec	99.05	96.87	76.01
1 Sec	92.14	86.23	80.57

TABLE IV
PERFORMANCE UNDER VARIOUS STRIDE LENGTHS AND FULLY CONNECTED LAYER COMPUTATION UNITS

Stride	Fully Connected Layer Settings			
	$(4096, 256)$	$(256, 128)$	(256)	(128)
3	99.73	98.66	98.92	98.37
2	99.05	98.29	96.98	97.32

We find that the acoustic data segment of 0.5 sec and the image size of dimension $224x224$ perform optimally with a max-pooling layer stride length of 2 and 128 computational units on the first fully connected layer. We assume that such parameter findings will be beneficial for the future real-time implementation. In the next section, we report the experimental findings based on the optimal settings in details.

V. RESULTS

We report our experimental findings on *eSense's* acoustic data in this section. Our experiments are mainly focused on the performance evaluation between different machine learning models and the application of domain adaptation techniques in the field of earable-based respiratory symptoms detection. Model performance is measured in terms of symptom detection accuracy.

[2]https://github.com/virufy/virufy-data

A. Model Performances Comparison

In comparing the traditional machine learning models with the deep learning model, we ensure that all machine learning models and the input features are at the optimum settings. Initially, we extract 46 statistical features and out of 46 features, we leverage the optimal number of features across the traditional models. Under the optimum performance settings, we carry out LOSO evaluation to measure the performances. Detailed model comparisons are presented in Figure 4.

The purple color bars represent the performance comparisons among different models and show that the deep learning model outperforms the traditional models by large margins. In order to evaluate the model scalability, we test the trained models on a different dataset and we observe that most of the traditional machine learning models do not deviate from the performance whereas the deep learning model performance degrades from 97.15% to 27.48%.

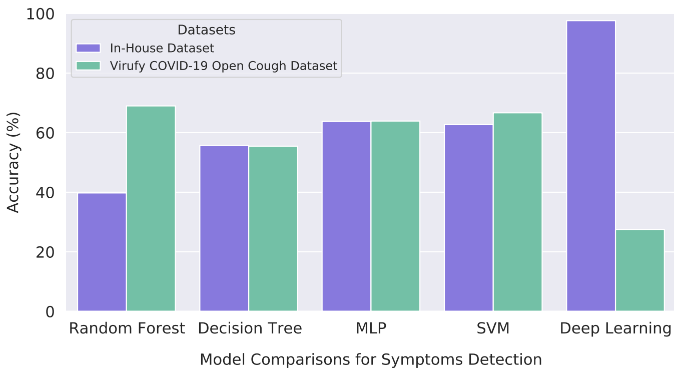

Fig. 4. Model Performances Comparison

B. Domain Adaptation Feasibility Study

Even though the reported model performances in different literature are quite high, our experiment reveals that the performance drops when a trained model is exposed to unseen data samples. It is mainly due to the data distribution heterogeneity caused by various factors such as device sampling frequency, device placement, user behaviour, and data collection procedure. We aim to increase the robustness of our trained deep model by implementing a domain adaptation technique. In the adaptation process, we use all three classes from our In-house dataset and the only available class (coughing) samples from the Virufy IV-A dataset. Our adaptation of the RevGrad layer [13] indeed recovers some of the performance loss, increasing the accuracy from 27.48% to 47.65%.

Apart from implementing a domain adaptation technique, we attempt to investigate the reasoning for such performance degradation through visual inspection of the dataset samples. Figure 5 shows a distinctive coughing pattern between two users which further causes different induced silence lengths in the audio files.

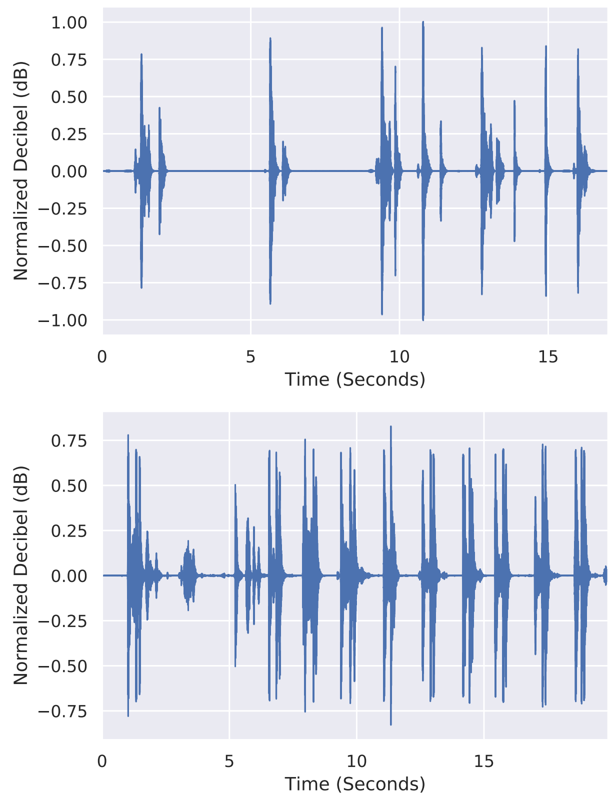

Fig. 5. Visual Inspection of Dataset Samples [Top: In-house Dataset (young male participant), Bottom: Virufy Dataset (elder female participant)]

VI. Discussion

We discuss a few issues that we have faced during our study ranging from dataset preprocessing to the prospective application of such proposed methodology in real-time systems.

- Data Processing: During our data preprocessing, we do not remove the silences. In many cases, silence is also considered as noise. Considering a hypothetical scenario where there could have been other noises instead of silence, such as dishwashing and car honking, removing the silence segments from the audio files will not make much of a difference if such noises are included. However, removing the silence and investigating the performance could be a future step in domain adaptation methodology.
- Real-time Implementation: Integrating the symptom detection pipeline into a smartphone is trivial but there are several implementation issues such as application power usage, application operation in the background, and detection of the non-symptom audio signals. These are several undeniable hindrances for real-time implementation.
- Mel-spectrogram Generation: In our deep learning methodology, we have generated Mel-spectrograms from the audio samples which might trigger an additional processing issue as well as device memory issue. Instead, if the audio data could be processed in its original form, such steps could be avoided. Thus, the domain adaptation

performance under such a situation could be one with the potential to work.

- Coughing Frequency Detection: Along with symptom detection, detecting the number of times a person suffers from any respiratory symptoms could also be a complementary measurement for symptom severity detection.

VII. Conclusion

We studied the feasibility of earable-provided acoustic data for respiratory symptom detection using various machine learning models. First, we collected a dataset for our feasibility study. We further extended our feasibility study once we found that when a trained model was tested with a dataset different from the training dataset, the performance degraded, providing a clear indication for the application of domain adaptation to increase the trained-model scalability. Our initial adaptation of an adversarial domain adaptation technique over a publicly available cough dataset reveals a promising sign for model scalability and performance improvement. In the future, we aim to investigate the issues discussed in the discussion section.

Acknowledgment

This research is funded by the National Science Foundation Grant CNS- 2050999.

The authors are thankful to all the Research Experience for Undergraduates (REU) program participants, program affiliated officials, and graduate student mentors for their valuable feedback in this research, especially to the volunteers who helped beyond scope during the data collection procedure in the pandemic.

References

[1] F. Kawsar, C. Min, A. Mathur, M. Van den Broeck, U. G. Acer, and C. Forlivesi, "esense: Earable platform for human sensing," in *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, pp. 541–541, 2018.

[2] J. Prakash, Z. Yang, Y.-L. Wei, and R. R. Choudhury, "Stear: Robust step counting from earables," in *Proceedings of the 1st International Workshop on Earable Computing*, pp. 36–41, 2019.

[3] T. Röddiger, D. Wolffram, D. Laubenstein, M. Budde, and M. Beigl, "Towards respiration rate monitoring using an in-ear headphone inertial measurement unit," in *Proceedings of the 1st International Workshop on Earable Computing*, pp. 48–53, 2019.

[4] E. Miller, N. Banerjee, and T. Zhu, "Smart homes that detect sneeze, cough, and face touching," *Smart Health*, p. 100170, 2020.

[5] J. Amoh and K. Odame, "Deepcough: A deep convolutional neural network in a wearable cough detection system," in *2015 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pp. 1–4, IEEE, 2015.

[6] R. X. A. Pramono, S. A. Imtiaz, and E. Rodriguez-Villegas, "Automatic cough detection in acoustic signal using spectral features," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 7153–7156, IEEE, 2019.

[7] T. Elfaramawy, C. L. Fall, M. Morissette, F. Lellouche, and B. Gosselin, "Wireless respiratory monitoring and coughing detection using a wearable patch sensor network," in *2017 15th IEEE International New Circuits and Systems Conference (NEWCAS)*, pp. 197–200, IEEE, 2017.

[8] I. D. Miranda, A. H. Diacon, and T. R. Niesler, "A comparative study of features for acoustic cough detection using deep architectures," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2601–2605, IEEE, 2019.

[9] J. E. Bardram, "The cams esense framework: Enabling earable computing for mhealth apps and digital phenotyping," in *Proceedings of the 1st International Workshop on Earable Computing*, pp. 3–7, 2019.

[10] K. Matsumura and K. Okada, "esense veers: A case study of acoustical manipulation in walking without sight both on subtle and overt conditions," in *Proceedings of the 1st International Workshop on Earable Computing*, pp. 20–25, 2019.

[11] M. Wang and W. Deng, "Deep visual domain adaptation: A survey," *Neurocomputing*, pp. 135–153, 2018.

[12] Y. Chang, A. Mathur, A. Isopoussu, J. Song, and F. Kawsar, "A systematic study of unsupervised domain adaptation for robust human-activity recognition," in *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, pp. 1–30, 2020.

[13] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *Proceedings of the 32nd International Conference on Machine Learning*, 2015.

[14] B. McFee, C. Raffel, D. Liang, D. P. Ellis, M. McVicar, E. Battenberg, and O. Nieto, "librosa: Audio and music signal analysis in python," in *Proceedings of the 14th python in science conference*, vol. 8, pp. 18–25, Citeseer, 2015.

[15] J. Lyons, D. Wang, H. S. Gianluca, E. Mavrinac, Y. Gaurkar, W. Watcharawisetkul, S. Birch, L. Zhihe, J. Hölzl, J. Lesinskis, *et al.*, "James lyons/python speech features: Release v0. 6.1," *Zenodo. doi*, vol. 10, 2020.

[16] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," 2017.

[17] 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*, pp. 3474–3484, 2020.

[18] M. Christ, N. Braun, J. Neuffer, and A. W. Kempa-Liehr, "Time series feature extraction on basis of scalable hypothesis tests (tsfresh–a python package)," *Neurocomputing*, vol. 307, pp. 72–77, 2018.

[19] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.