Efficient Online Cough Detection with a Minimal Feature Set Using Smartphones for Automated Assessment of Pulmonary Patients

Md Juber Rahman
Electrical and Computer Engineering Department
The University of Memphis
Memphis, USA
e-mail: mrahman8@memphis.edu

Ebrahim Nemati, Mahbubur Rahman, Korosh Vatanparvar, Viswam Nathan, Jilong Kuang
Digital Health Lab
Samsung Research America
Mountain View, USA
e-mail: e.nemati@samsung.com

***Abstract*—An automated monitoring of chronic diseases may help in the early identification of exacerbation, reduction of healthcare expenditure, as well as improve patient's health-related quality of life. Cough monitoring provides valuable information in the assessment of asthma and Chronic Obstructive Pulmonary Disease (COPD). In this multi-cohort study, we have used every-day wearables such as smartphone and smartwatch to collect cough instances from 131 subjects including 69 asthma patients, 9 COPD patients, 13 patients with a co-morbidity of asthma and COPD and 40 healthy controls. For online cough detection we have identified the audio features suitable for resource-constrained platforms (e.g., smartphone), ranked the features and identified top 9 features to obtain an F-1 score of 99.8% in the offline classification of 23,884 cough instances from non-cough (speech/silence, etc.) events using Random Forest classifier. Finally, a power and time-efficient scheme for continuous online cough detection from the audio stream has been illustrated. The proposed model has an online cough detection sensitivity of 93.3%, specificity of 98.8% and accuracy of 98.8%. In addition, a good improvement in reducing the on-device execution (feature extraction and classification) time and power consumption has been achieved compared to the current state of the art algorithms. The proposed on-device cough detector has been implemented to meet the criteria for integration in the passive monitoring and online assessment of asthma/COPD patients.**

Keywords- cough; online detection; pulmonary disease; random forest; streaming audio.

I. INTRODUCTION

Lung diseases are among the biggest killers in the world. In the USA, lung disease is the third leading cause of death [1][2]. Many of the lung diseases are chronic conditions in nature which severely impacts the patients' health-related quality of life. As a result, the associated healthcare expenditure is substantial. The annual direct and indirect healthcare cost related to obstructive lung diseases such as asthma and COPD has been estimated to be $154 billion [3]. Early diagnosis and follow-up have the potential to reduce hospital visits, associated expenditures, and improve patients' quality of life.

Spirometry and standard questionnaires have been used extensively in the diagnosis and severity estimation of asthma and COPD. Monitoring of warning signs such as cough, shortness of breath, etc. has been proven to be useful in the detection and management of asthma and COPD [4]. Usually, cough frequency and severity are reported by the patient himself. This approach is highly subjective and not suitable for continuous passive monitoring. As an alternative, there have been attempts in developing automated cough monitors.

Audio signal from the acoustic sensor has been primarily used as the basis for automatic cough detection. Though there are multi-sensor approaches that include non-acoustic sensors for automatic cough detection, previous research efforts indicate that high sensitivity and accuracy can be achieved with audio signal solely [5]. Also, employing acoustic sensor seems to be the most suitable for 24 h ambulatory monitoring. Commonly used features for cough detection include audio spectral features, Mel-Frequency Cepstral Coefficients (MFCC), Linear Prediction Cepstral Coefficients (LPCC), Hu moments, etc. Birring *et al.* introduced an automatic cough detection system called Leicester cough monitor using Hidden Markov Model (HMM) [6]. The system incorporates 24 h ambulatory recording solely relying on the acoustic signal. They obtained a sensitivity of 91% and specificity of 99% with spectral audio features. Larson *et al.* proposed a low-cost microphone-based cough sensing system using Random forest classifier [7]. These approaches requiring a specialized device incur extra cost and burden for the user as he needs to carry that extra device all the time. Shin *et al.* investigated a hybrid model consisting of both Artificial Neural Network and HMM [8]. They used sound pressure level, cepstral coefficient, and temporal features of audio and obtained 91.3% accuracy for cough detection. However, their dataset is too small containing only 143 cough sounds and 110 environmental sounds. Also, it is based on MATLAB, which is not suitable for on-device detection. Liu *et al.* investigated a combination of deep neural network and HMM for automatic cough detection using MFCC features [9]. Amoh *et al.* investigated the use of a convolutional neural network in a wearable cough detection system [10]. It is noteworthy from these works that while deep learning imposed a high computational cost, there was no significant improvement in classification performance compared to traditional methods.

Cough detection from recorded audio is subjected to privacy compromise and hence has never been popular or widely accepted. Recent advancement in the quality of acoustic sensors and processing capacity of smartphones and wearable devices has triggered a growing tendency for online cough detection from streaming audio without recording the audio.

 ISBN: 978-1-61208-739-9

While automatic cough detection is well investigated, few studies addressed on device feature extraction and classification from streaming audio. Pham et al. investigated a Gaussian mixture model and a universal model for real-time cough detection using smartphones [11]. They achieved a sensitivity of 81% for subject independent training, however, they did not address system performance-related issues. Most recently, J. Alvarez et al. investigated the efficient computation of image moments for robust cough detection using smartphones [12]. While they achieved 88% of sensitivity for cough detection the app power consumption is 25% of the device power consumption for 24 hours of usage. Also, the time required for feature extraction is relatively long (5 min 28 secs, as it requires image processing) and needs to be reduced for efficient online implementation. Another limitation of previous studies is their inability to discriminate between the cough of the intended subject with the cough of other subjects, which make the cough monitoring ineffective in a social or family setting if multiple people have cough syndrome. E. C. Larson et al. reviewed the shortcomings of existing cough detection approaches in details and described the need for further investigation for smartphone-centric ambient audio sensing for effective pulmonary assessment [13]. mLung Study is our comprehensive initiative aimed to leverage the power of wearables and smartphones for early detection and continuous monitoring of asthma and COPD patients which include quality data collection, multi-layer annotation, on-device feature extraction and classification, privacy protected in-depth analysis in the cloud, etc. Previously, we reported a framework for maintaining privacy while recording audio for offline cough classification [14]. In this paper, we report a model implemented and tested on android smartphones for online cough detection from streaming audio. The model was trained on a large dataset containing both voluntary and natural coughs. Contributions of this study have been summarized below:

i) Identification of audio features, optimal window size and overlapping suitable for automatic cough detection with high sensitivity using resource-constrained devices such as a smartphone.

ii) Feature ranking and classifier optimization for computational efficiency to minimize the execution time while retaining classification performance.

iii) Enabling subject-specific cough detection and discriminating secondary subject coughs.

iv) Analysis and optimization of system overhead to achieve better performance for online processing in smartphones.

This study presents promising results for using smartphones in a privacy-preserved personalized and reliable online cough detection framework which facilitates the online assessment of asthma and COPD patients as well as healthy population.

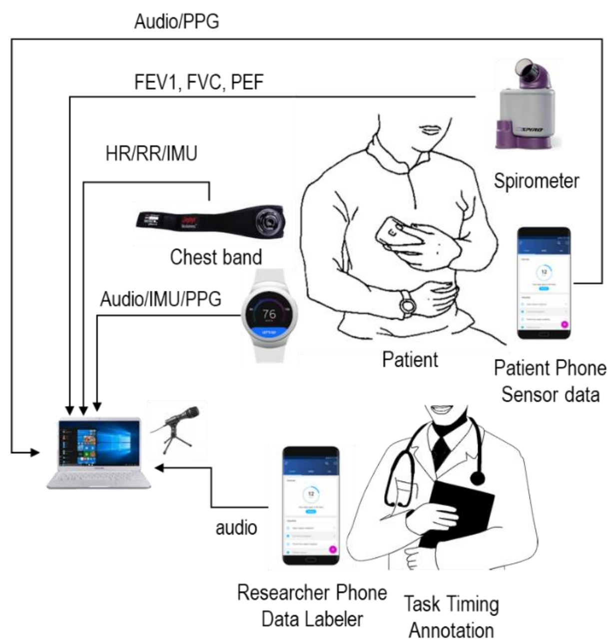

Figure 1 Study description and cough recording protocol

The remainder of this paper is organized as follows. Section 2 describes the materials and methods including the online cough detection framework and the process of system performance evaluation. Section 3 describes the result of the off-line analysis, on-line cough detection performance, and system performance for real-time processing. Finally, Section 4 presents our conclusion and future work scope.

II. MATERIALS AND METHODS

This section describes the dataset, study protocol, algorithm and framework in details:

A. *Description of Subjects and Study Protocol*

Per institutional review board (IRB) approval, a total of 131 subjects (67 males and 64 female) were recruited for this study out of which 40 were healthy controls without any diagnosed medical condition and 91 individuals were suffering from pulmonary diseases. All of the patients have been diagnosed with pulmonary diseases by medical practitioners. Out of the 91 patients 69 were diagnosed with asthma, 9 with COPD and 13 exhibited a co-morbidity of asthma and COPD.

The subjects were from different racial backgrounds including African-American, Asian, Caucasian, and Native American and had an age range from 14 to 82 years. During the recruitment process, all subjects with the history of cardiac disease i.e. arrhythmia or heart attack, pulmonary infection, vocal cord dysfunction, and inability to read or speak English were excluded. After obtaining informed written consents, spirometry-based pulmonary function test was done for all of the patients. Multimodal physiological data such as ECG, PPG, audio, and IMU were collected from the subjects in a laboratory setup using multiple wearable sensors including smartwatch (Samsung Gear Sport), chest band (Zephyr BioHarness 3.0, Medtronic plc), and smartphone (Samsung Galaxy Note 8).

 ISBN: 978-1-61208-739-9

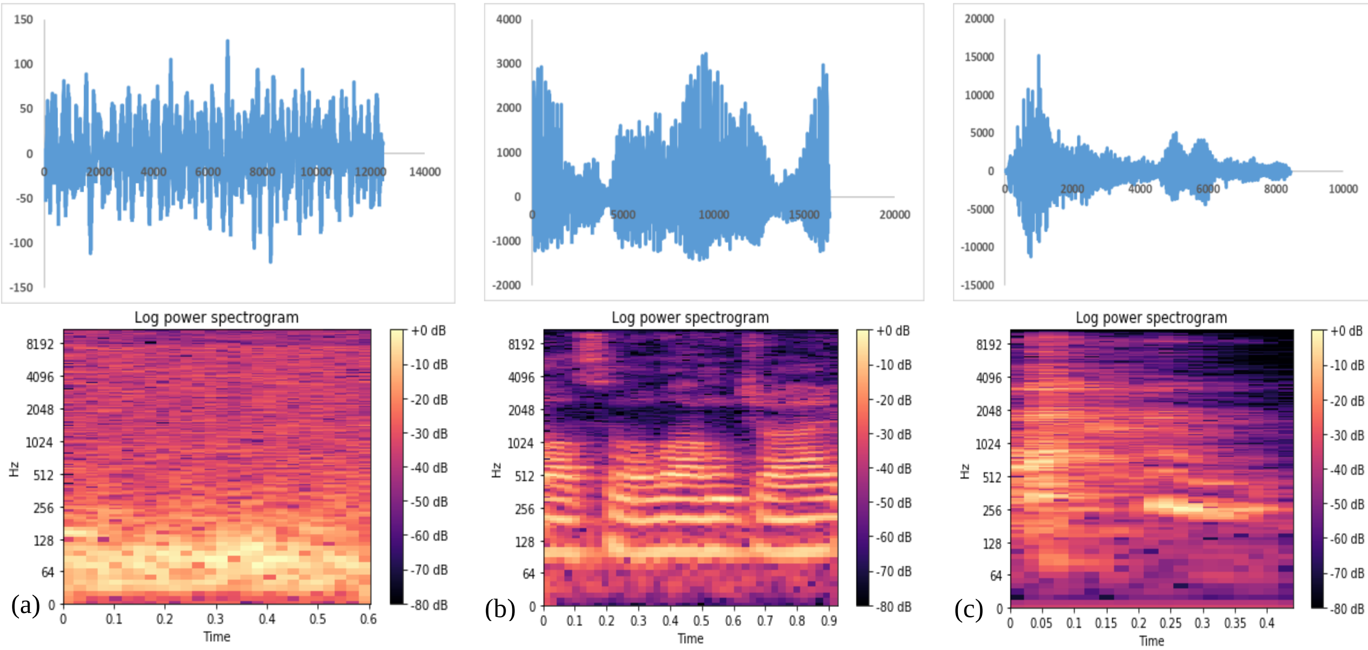

Figure 2 Typical (a) silent, (b) speech, and (c) cough episodes and their spectral representations showing the differences in frequency component and loudness.

The length of the data collection session was 30-40 minutes. During this time tasks related to the pulmonary patient assessment, were performed which included PFT at the beginning and the end. The audio was captured from both smartphone and smartwatch with a sampling rate of 44.1kHz. Participants wore a Samsung Gear S3 smartwatch on their left hand. They held the smartphone (Samsung Galaxy Note 8) on the left side of the chest to capture chest motions as well as lung sounds such as wheezes. The system architecture utilized for the data collection can be seen in Figure 1. The experiment protocol that is specifically designed for asthma and COPD patients included the following sessions:

- Pulmonary Function Test-1: standard mobile spirometry.
- Sit-Silent Breathing: sitting silently for one minute and counting breaths while keeping the phone on the chest and watch on the abdomen.
- Supine-Silent Breathing: repeat the previous task in the supine position.
- Cough: produce several voluntary coughs for up to two minutes. Also, counted natural coughs.
- A-vowel Voice: vocalizing '*Aaaa*....' sound for as long as they can.
- Speech: speaking freely about any topic of interest.
- Reading: read aloud a standard passage.
- Pulmonary Function Test-2: standard mobile spirometry.

The rationale behind the study protocol has been described earlier [15].

B. Audio Processing, Data Preparation and Feature Extraction

The entire record audio has been annotated manually for cough, speech, and silence by a crowdsourcing annotation platform, FigureEight [16]. For cough detection purpose, wheezes and other body sounds have been included in the speech category. In addition to recorded audio, spectral visualization of an audio signal has been used in the annotation process to improve the quality of annotation. Figure 2 shows the time-domain and corresponding spectral representation of silent, speech and cough episodes. The cough instances are characterized by a burst followed by a voiced part which makes them distinguishable from the speech and silence. From the spectrogram, it is clear that frequency components and loudness of the cough are very different from that of regular speech or silence. The start and stop time of each cough event has been marked in the annotation process and then the recording has been segmented into cough, speech and silent episodes and labeled accordingly. Finally, 23884 cough instances, 165948 speech instances, and 52135 silent episodes were obtained from the audio clips. For subject discriminatory cough detection, 35 coughs from one subject have been placed in one class while the other class had 170 cough instances from multiple subjects. Each wav file is a 16-bit PCM-encoded audio with the sampling rate of 44100 samples/sec.

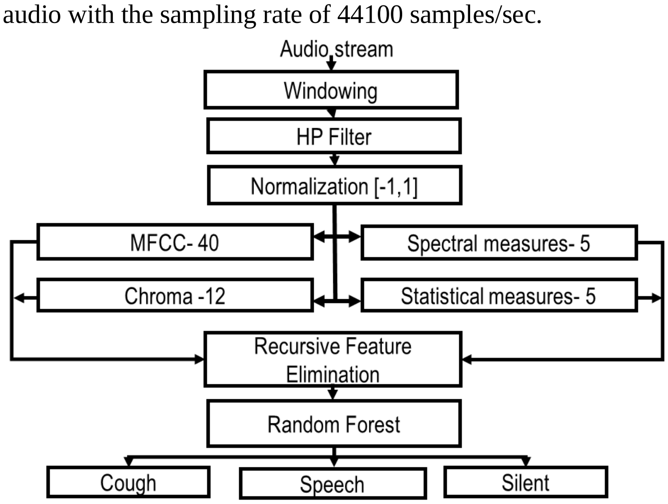

Figure 3 Method for feature extraction, feature selection, and classification.

For feature extraction, the wav file was chopped into frames of 0.6 sec, high pass filtered (200 Hz) and normalized in the range [-1,1]. An overlapping of 10% has been used between the frames for online implementation. Features were then extracted from the frames which included time-domain features such as absolute mean, absolute median, standard deviation, skewness, kurtosis, zero-crossing rate and frequency features such as spectral centroid, spectral roll-off, spectral variance, MFCC, and spectral chroma. The feature set also includes energy and sound pressure level. An open-source library, Taros-DSP, has been used to read the audio from microphone, process the audio file and extract MFCC features [17]. Another open-source library jMusic has been used for extracting the spectral features [18]. Signal pre-processing and feature extraction were done in JAVA. The process of feature extraction, feature selection and classification has been shown in Figure 3. To compute the MFCC features, Fast Fourier Transform with a hamming window has been used in estimating the magnitude spectrum. The number of Mel filters used is 50, the lower filter frequency is 300 Hz and the upper filter frequency is 8000 Hz [19].

C. Sound Event Detection

Sound events can be detected based on different features. In this work, we have used sound pressure level and energy to detect the sound event. The mean sound pressure level of all silent episodes has been used as the threshold for sound event detection. Any episode with a sound pressure level greater than mean value has been considered as a sound event followed by classification performed to detect if it is a cough, speech or silent episode. The possibility of missing a sound event has been almost eliminated due to this dynamic thresholding. This makes the cough detection feasible even when the smartphone is relatively far from the patient. Audio episodes with mean less than the threshold, are not considered as an event and therefore skipped for feature extraction and classification which helped with the overall power consumption.

D. Feature Selection, Classification, and offline-Training

Dimension reduction is important to reduce the time and computational complexity associated with the implementation of algorithms on wearables. Also, optimal feature selection is an important step to enhance the performance of classification and ensure better generalization. In this work, we have used the recursive feature elimination technique for selecting the top-ranked features. Caret package from R has been used for the feature ranking [20].

For finding the best classification model we have explored logistic regression, support vector machines (SVM) with different kernels and random forest. The decision tree has been used as the base classifier for random forest and samples are drawn with replacement. The number of estimators used in the random forest is 100. To reduce memory consumption and the time complexity, maximum depth (=20) of the tree has been decided using a heuristic approach. 10-fold cross-validation was employed on the training data to evaluate classifier performance and adju st the hyper-parameters. WEKA has been used as the model development environment [21]. The subject discriminatory model has been trained and evaluated separately.

E. Online Cough Detection and Counting Framework

i) Design Goals and Considerations

- Passive sensing- no user effort is expected.
- Privacy-preserving- since speech and related data can reveal user identity, no audio is being recorded. Classification is done using the features generated on the device.
- Reliable detection- high sensitivity not to miss any cough instances.
- Execution Time- keep the processing time for feature extraction and classification low enough to facilitate real-time processing and prediction.
- Performance Optimization- the high emphasis has been given to keep the app power consumption, memory usage and latency low, so that device normal functionality is not drastically impacted by the app.

ii) System Overview, Implementation, and Evaluation

The online cough detection framework has been shown in. Figure 4.

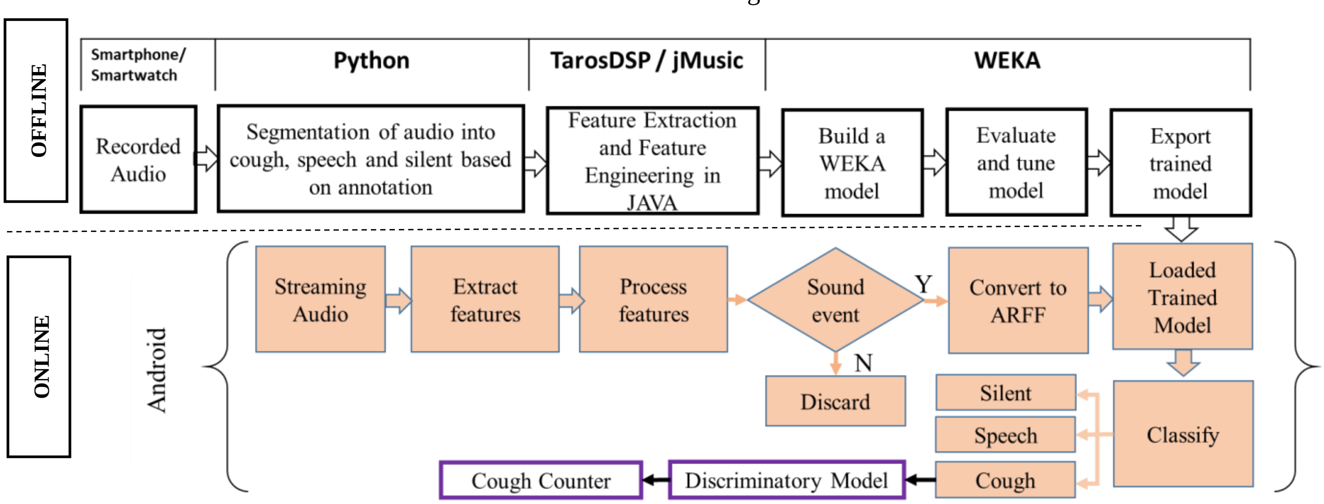

Figure 4 Online cough counter framework

Trained, evaluated and tuned model from WEKA has been exported for use in Android. In android, the trained model was stored in the asset directory and was loaded in the activity for online classification of audio frames using on-device extracted features. The audio signal was directly read from the microphone in an audio buffer and was processed as an audio event in 0.6 sec frames and prediction is made for each of these frames. Confirmed silent episodes are discarded and no further feature extraction/classification is done to reduce execution time and power consumption. The extracted features, as well as the predictions, are written to a CSV file and exported to the cloud for in-depth offline processing. For evaluating the online cough detection performance, the app has been tested for 2 days in the real-world scenario which include home environments, driving, walking in the street and social gathering.

III. Results

The boxplots for the sound pressure level of cough, speech and silent episodes have been shown in Figure 5. It is evident that the sound pressure level of silent episodes is much lower compared to cough and speech. Figure6 shows the result of feature ranking by recursive feature elimination technique. The feature ranking suggests that best performance (good accuracy and low dimension) can be achieved with 9 top-ranked features. The top-ranked features are *mfcc_0*, *pressure level*, *standard deviation*, *kurtosis*, *mean*, *mfcc_1*, *median*, *zero-crossing rate*, and *mfcc_2*. Figure 7 shows the boxplot comparison between cough and speech events for the top-ranked MFCC features. A good visual separation between cough and speech episodes can be observed in the boxplot comparison. The classification performance of different classifiers with the top-ranked features has been shown in Table I. Random Forest performed best with a precision of 99.8%, recall of 99.8% and an F-1 score of 99.8% for 10-fold cross-validation. The confusion matrix for 10-fold cross-validation has been shown in Table II. Only 4 cough instances have been misclassified out of 23884 cough instances. Figure 8 shows the model build time, test time and F-1 score at different depths of the forest for the Random Forest classifier. Low model build time is important for subject-specific cough detection as it requires online training. It can be seen that increasing the depth beyond 20 increases the build and test time with a minimal gain in F-1 score. Hence, the optimal depth is found to be 20 to create the final model. A precision of 94.2%, recall of 94.1% and F-1 score of 93.7% have been achieved with Random Forest in detecting cough from the intended subject while discriminating coughs from other subjects as shown in Table III.

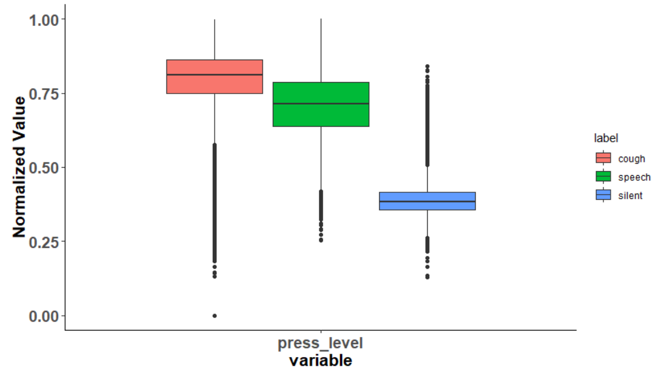

Figure 5 Boxplots showing normalized sound pressure level for cough, speech and silence instances

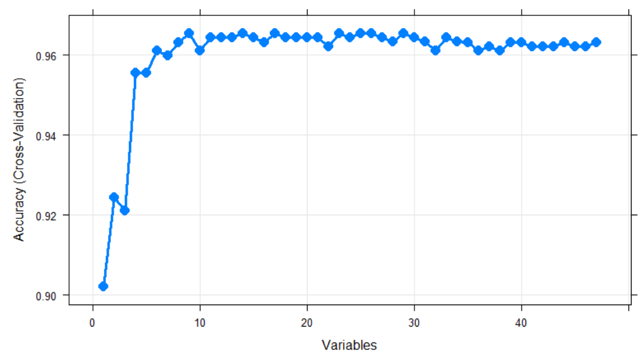

Figure 6 Optimal no. of features using recursive feature elimination technique

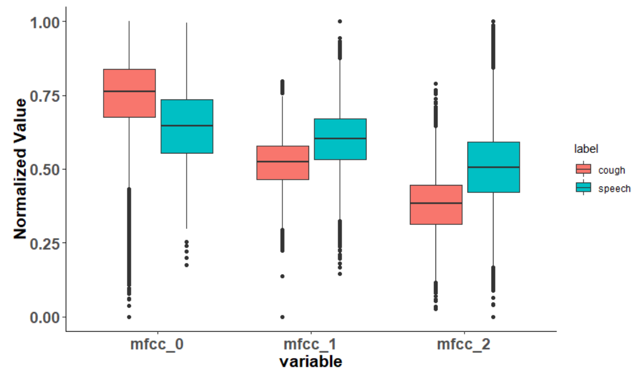

Figure 7 Boxplot comparison between cough and speech for top-ranked MFCC features

TABLE I. Off-line classification performance with different classifiers (10-fold CV)

Classifier	Cough detection		
	precision	***recall***	***F-1 score***
Logistic Regression	93.0%	92.9%	92.9%
SVM (kernel=Poly)	93.1%	93.0%	93%
Random Forest	**99.8%**	**99.8%**	**99.8%**

TABLE II. CONFUSION MATRIX FOR RANDOM FOREST CLASSIFIER (10-FOLD CV)

cough	*speech*	*silence*	
23880	4	0	*cough*
0	165944	468	*speech*
0	54	52081	*silence*

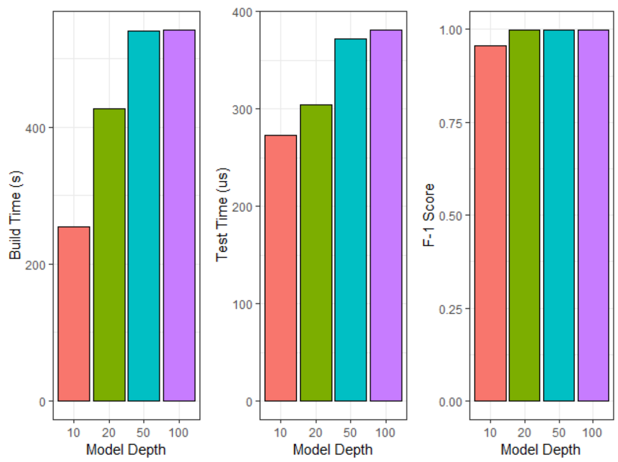

Figure 8 Build time, test time and F-1 score at different depths of the Random forest classifier

TABLE III. CLASSIFICATION PERFORMANCE FOR SUBJECT-SPECIFIC COUGH DETECTION (10-FOLD CV)

Classifier	**Cough detection**		
	precision	***recall***	***F-1 score***
Logistic Regression	89.2%	88.8%	89.0%
SVM	93.3%	93.2%	93.2%
Random Forest	94.2%	94.1%	93.7%

TABLE IV. SYSTEM OVERHEAD FOR ON-DEVICE FEATURE EXTRACTION AND COUGH CLASSIFICATION IN SMARTPHONES FROM STREAMING AUDIO

App	***Latency***	***Memory***	***Avg. CPU***	***Power Consumption***
Cough Counter	375 ms	99 MB	8.69 %	55 mAh

Using the implemented model, a sensitivity of 93.3%, specificity of 98.8% and accuracy of 98.8% have been achieved for online cough detection. The feature extraction and classification time for a 2 min audio clip is only 9.8 secs which is much lower compared to other feature set reported in previous studies [12]. The system overhead on a smartphone for running the cough detection app continuously has been shown in Table IV and compared with VoiceOver app (already available in the play store, 500K+ downloads) in Figure 9. The functionality of VoiceOver app includes audio recording, audio processing, sharing and audio storage; whereas the functionality of Cough app includes audio sampling, audio processing, feature extraction and classification and export/storage of feature values. It can

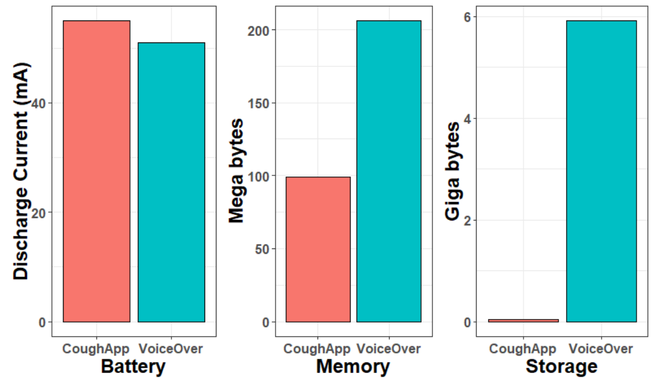

Figure 9 Comparison of the performance metrics of Cough Counter app with VoiceOver app

TABLE V. COMPARISON OF THIS WORK WITH PREVIOUS STUDIES

Ref.	Platform	Subjects (Healthy/ Patient)	Online cough detection	
			Classifier	Performance
[6]	Specialized Wearable	71 (8/65)	HMM	Sen. 91 % Sp. 99%
[8]	Specialized Wearable	84	ANN+ HMM	Sen. 91.3%
[10]	Specialized Wearable	14 (14/0)	CNN	Sen. 95.1% Sp. 99.5%
[11]	Smartphone	Not mentioned	GMM UBM	Sen. 91%
[12]	Smartphone	13 (0/14)	kNN	Sen. 88.5% Sp. 99.77%
[7]	Smartphone	17 (0/3), other-14	RF	TPR-92% FPR-0.5%
Proposed Work	Smartphone	131 (40/71)	RF	Sen, 94.3% Sp. 98.8%

be observed that storing feature values instead of audio require much lower storage space. The Cough app consumes less memory but more power than VoiceOver app. Nonetheless, the power consumption (11% of the device total power usage) is lower than previously reported (25% of the device total power usage) cough detection framework [12]. Minimal no. of features, optimal forest depth and silence removal (processing less no. of frames) have contributed to reducing the power consumption. The low latency and CPU usage of cough app are suitable for continuous operation. All testing has been performed using a Samsung Galaxy Note 8 smartphone. A comparison of this work with similar previous studies has been shown in Table V. It can be seen that other smartphone based approaches for cough detection have much lower performance compared to the proposed method [7] [11] [12]. In addition, the size of their dataset is very small which will impact the generalization capacity of the developed models.

IV. CONCLUSION

Cough pattern analysis may be helpful in monitoring asthma and COPD patients passively. However, the privacy of the users is at great risk when it comes to continuous listening if the processing has to be done on the cloud. We

 ISBN: 978-1-61208-739-9

have proposed an on-device cough detection framework that detects the cough occurrence from the streaming audio without the need to store the audio on the device or send it to the cloud. To reduce the computational burden, we have ranked the features and identified the top 9 features to obtain a reasonable accuracy and optimized the classifier to have low complexity while providing a high accuracy of 98.8%. This approach is computationally efficient and suitable for smartphones. Our future work includes the improvement of model generalization performance and robustness. Also, we are planning to implement this cough detector as a module among other modules to provide an assessment of the severity of asthma and COPD patients.

REFERENCES

[1] Global Initiative for Chronic Obstructive Lung Disease (GOLD) 2019 "Global Strategy for the Diagnosis, Management and Prevention of COPD" Available from http://www.goldcopd.org Accessed May 20, 2019.

[2] K. D. Kochanek, S. L. Murphy, J. Q. Xu , and B. Tejada-Vera, "Deaths: Final data for 2014." *National vital statistics reports*; vol 65, no 4. pp. 1-122, Jun 2016.

[3] Lung Institute, "The Cost of Lung Disease" Available: https://lunginstitute.com/blog/the-cost-of-lung-disease/, Accessed May 20, 2019.

[4] E. R. McFadden Jr, "Clinical physiologic correlates in asthma." *Journal of allergy and clinical immunology* 77, no. 1, pp. 1-5, 1986.

[5] T. Drugman *et al*., "Objective Study of Sensor Relevance for Automatic Cough Detection," in *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 3, pp. 699-707, May 2013.

[6] S.S.Birring *et al*. ,"The Leicester cough monitor: preliminary validation of an automated cough detection system in chronic cough," Eur. Respiratory J., vol. 31, no. 5, pp. 1013–1018, May 2008.

[7] E. C. Larson, T. J. Lee, S. Liu, M. Rosenfeld, and S.N. Patel. "Accurate and privacy preserving cough sensing using a low-cost microphone." In *Proceedings of the 13th international conf. on Ubiquitous computing*, pp. 375-384. ACM, 2011.

[8] S. Shin, T. Hashimoto, and S. Hatano, "Automatic Detection System for Cough Sounds as a Symptom of Abnormal Health Condition," in *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 4, pp. 486-493, July 2009.

[9] J. M. Liu *et al.*,"Cough event classification by pretrained deep neural network." *BMC medical informatics and decision making* vol. 15 Suppl 4, 2015.

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

[11] C. Pham, "MobiCough: real-time cough detection and monitoring using low-cost mobile devices." *Asian Conference on Intelligent Information and Database Systems*. Springer, Berlin, Heidelberg, 2016.

[12] J. Monge-Álvarez and C. Hoyos-Barceló, "Robust Detection of Audio-Cough Events Using Local Hu Moments," in *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 1, pp. 184-196, Jan. 2019.

[13] E. C. Larson, E. Saba, S. Kaiser, M. Goel, and S. N. Patel. "Pulmonary Monitoring Using Smartphones." In *Mobile Health*, pp. 239-264. Springer, Cham, 2017.

[14] E. Nemati *et al.*, "Private Audio-Based Cough Sensing for In-Home Pulmonary Assessment using Mobile Devices" *13th International Conference on Body Area Networks*, 2018.

[15] M. Rahman *et al*., "Towards Reliable Data Collection and Annotation to Extract Pulmonary Digital Biomarkers Using Mobile Sensors" *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare*. ACM, 2019.

[16] https://www.figure-eight.com/, accessed on May 1, 2019.

[17] J. Six, O. Cornelis, and M. Leman, "TarsosDSP, a real-time audio processing framework in Java." *Audio Engineering Society Conference: 53rd International Conference: Semantic Audio*. Audio Engineering Society, 2014.

[18] A. R. Brown and A. C. Sorensen, "Introducing jmusic." *InterFACES: Proceedings of The Australasian Computer Music Conference.* Brisbane: ACMA, pp. 68-76, 2000.

[19] X. Huang, A. Acero, and H. Hon, *Spoken Language Processing: A guide to theory, algorithm, and system development*. Prentice Hall, 2001.

[20] M. Kuhn, "Building predictive models in R using the caret package." *Journal of statistical software* 28, no. 5 ,pp.1-26, 2008.

[21] M. Hall *et al*. "The WEKA data mining software: an update." *ACM SIGKDD explorations newsletter* 11, no. 1, pp.10-18, 2009.