TOWARDS A WEARABLE COUGH DETECTOR BASED ON NEURAL NETWORKS

¹Prad Kadambi, ¹Abinash Mohanty, ¹Hao Ren, ²Jaclyn Smith, ²Kevin McGuinnes, ²Kimberly Holt, ³Armin Furtwaengler, ³Roberto Slepetys, ³Zheng Yang, ¹Jae-sun Seo, ¹Junseok Chae, ¹Yu Cao, ¹Visar Berisha

Arizona State University
 ² University of Manchester
 ³ Boehringer Ingelheim

ABSTRACT

Persistent cough is a symptom common to a number of respiratory disorders; however, reliable monitoring of cough frequency and cough severity over an extended period of time can be a challenge. Traditional methods involve subjective evaluation by care providers or patient self-reports. As an alternative, we propose an objective method for monitoring cough using a wearable microphone. We collected 24-hour audio recordings from 9 patients suffering from chronic obstructive pulmonary disease, asthma, and lung cancer using the VitaloJAK wearable microphone. Trained professionals carefully listened to each audio stream and manually labeled each cough event. Using this data, we propose a new neuralnetwork-based cough detection scheme. A pre-processing algorithm is used to estimate the start and end of each cough and the deep neural network is trained using each cough instance. Experiments demonstrate an average leave-one-participantout cross-validation specificity and sensitivity of 93.7% and 97.6% respectively.

Index Terms— cough detection, deep learning, mobile health sensing, respiratory disease, audio processing

1. INTRODUCTION

Coughing is one of the most important and frequent symptoms reported by patients [1] [2]. Chronic cough can result in deleterious effects on health and quality of life [3]. Monitoring cough symptoms is important in detecting and treating respiratory conditions such as chronic obstructive pulmonary disease (COPD), asthma, pulmonary fibrosis, and tuberculosis [4] [5] [6].

To assess the frequency and severity of cough, several subjective tests have been developed (e.g. Leicester cough questionnaire [7], visual analog scales, etc.). These methods provide insight into the perceived severity of cough symptoms, but are ultimately unreliable when compared to objective methods of studying cough, because factors such as patient mood, vigilance, and the placebo effect can impact the patient's report of cough frequency [8] [9].

However, objective tools for studying cough are lacking. One quantitative method to assess cough frequency and severity consists of using ambulatory systems to record audio from patients for an extended time, and then manually counting the number of coughs in the recorded audio. Manually counting coughs is a time-consuming process that requires an expert to verify labeled coughs [10]. This is impractical for large amounts of data.

A number of automatic cough detection systems have been proposed in the literature. The Leicester Cough Monitor (LCM) [11], and VitaloJAK [12] are examples of ambulatory systems consisting of both wearable devices to record patient audio, and algorithms for cough detection from recorded data. The LCM applies a Hidden Markov Model (HMM) trained on mel-frequency cepstral coefficients (MFCCs) in order to detect cough sounds. However, the LCM algorithm is only semi-automated - it requires manual tuning of model parameters for each individual recording [13]. This algorithm takes a 24-hour patient audio recording and creates a shorter recording with all suspected coughs. To decrease false alarm rate, a portion of the detected coughs must be manually confirmed by personnel [11].

Recently, there has been significant interest in applying deep learning techniques to automatic cough detection. In [14] the authors devise a probabilistic neural network trained using linear predictive cepstral coefficients (LPCCs) and MFCCs to distinguish cough sounds from background. The authors in [15] and [16] applied convolutional neural networks (CNNs) trained directly on the short-time Fourier transform (STFT) of audio segments. However, most of these methods are validated on limited datasets collected in artificial environments. or use proprietary hardware for collecting patient data. For example, in [17] the data only consists of three patients recorded in a hospital setting; and, only one patient was recorded for more than four hours. The authors in [15] and [16] use custom hardware to record healthy volunteers reading passages and voluntarily coughing in a controlled lab setting. It is wellknown in the literature that voluntary coughs have different



Fig. 1. Spectrogram of speech and noise compared with three coughs

patterns from reflex coughs [18]. In [19] a pre-trained neural network is applied to 24-hour patient recordings collected in a restricted, hospital environment. They use custom recording devices rather than using an FDA-cleared cough monitor.

In this work, we propose a framework for audio-based automatic cough detection. The main contributions of this work are: (1) an extensive dataset containing 9 days of audio recorded in real-world conditions, from 9 patients with a variety of respiratory illnesses, using the FDA-cleared cough monitoring device, VitaloJAK; (2) a pre-processing algorithm to fine tune data labels to improve neural network accuracy and convert event based cough labeling to labels containing cough start and end points; (3) a deep neural network (DNN) trained using MFCCs and other features to discriminate cough sounds from background noise. The proposed framework achieves an average leave-one-out cross-validation specificity, sensitivity, and accuracy of 93.7%, 97.6% and 92.3% respectively.

2. METHODS

2.1. Data Collection

Recordings were supplied from an acoustic cough recording repository (RADAR) maintained at the University Hospital of South Manchester, with patient consent. Sound recordings were collected using the VitaloJAK cough recording device over 24-hour periods; recordings were commenced in a research clinic and then patients were permitted to go about their normal daily routines. The monitors were collected once

Participant	Disease	Coughs	Gender
1	Chronic Cough	3133	М
2	Chronic Cough	509	F
3	Chronic Cough	546	F
4	COPD	102	F
5	COPD	852	F
6	Asthma	221	F
7	Asthma	118	М
8	Lung Cancer	163	F
9	Lung Cancer	26	М

Table 1. Detailed Participant Information

the recordings were completed. The device makes continuous sound recordings at 8 kHz sample rate, from an air-coupled contact microphone placed over the manubrium sterni and a free-field lapel microphone. We use audio from the lapel microphone for our analysis. Participants were instructed not to remove the device or microphones during the recording and to keep the equipment dry. A total of 9 recordings (3 chronic cough, 2 asthma, 2 chronic obstructive pulmonary disease and 2 lung cancer patients) were included in the analysis. Each 24 hour recording was listened to in its entirety by technical staff trained in cough identification, and the location of each cough sound heard was recorded electronically.

Our dataset consists of a total of 5,670 coughs. The dataset contains a rich variety of background noise such as music, conversation, watching television, and riding in a car. The audio contains many sounds easily confused with coughing such as throat clearing, sneezing, and laughing. Table 1 shows the breakdown of the dataset by disease, cough count, and gender.

Figure 1, shows the spectrogram of three coughs of different length and a set of non-cough sounds taken from the dataset. We can make two important observations from this spectrogram. We note that coughing contains a larger amount of energy in higher frequencies than speech or other types of noise. A properly trained DNN can discriminate coughing from background by utilizing these characteristics unique to coughing. Also, any algorithm trained to detect these coughs must be able to account for the variability in cough length and intensity (see Figure 1).

2.2. Data Preprocessing

Every cough in the database was manually labeled by a trained expert. The top graph in Figure 2 shows how coughs were labeled. As we have noted, coughs vary in duration. However, the provided labels do not reflect this information. If features are extracted from a constant window around the provided labels, background audio events adjacent to coughing can be unintentionally included as part of the cough. Therefore, we must determine the cough start and end times from the provided labels.

The cough reflex consists of three audible portions: 1) a rapid, explosive phase, 2) an intermediate, decaying phase consistent with forced expiration, and 3) a voiced phase (not necessarily present in all coughs). Since the first two phases are ubiquitous across all coughs, they allow us to determine the start and end of a cough using an energy-based criteria.

Figure 2 outlines the label preprocessing algorithm. Given the event-based label, we extract a 420 ms window of audio from 70 ms before to 350 ms after the provided label. We choose a window of 420 ms because more than 95% of all coughs in our dataset were observed to be shorter than 400 ms in duration.

Next, an energy versus time profile is generated for the cough. We calculate the energy for every 10 ms frame within



Fig. 2. Preprocessing algorithm extracts a more descriptive cough label. Given a labeled cough (shown in top graph), we extract an energy vs. time profile for the cough (shown in middle graph), and use this to determine the cough duration

the 420 ms window using a step size of 2 ms. We calculate the energy for each 10 ms frame and find the maximum value of this energy profile within the 420 ms window. The first 10 ms frame that precedes the maximum energy frame with 15% of the energy of the maximum energy frame is chosen as the start of the cough; and the first frame that occurs after the maximum energy frame is selected as the end of the cough. Any cough found to have a duration of less than 40 ms is pruned from the dataset¹. The resulting dataset consists of coughs ranging from 40 ms to 420 ms duration, with an average duration of 200 ms.

2.3. Feature Extraction

A total of 168 features are used as inputs to the DNN. Since we aim to apply the DNN in a real-time setting in subsequent work, we perform training and inference using 200 ms frames of audio (200 ms corresponds to the average cough length). Four 200 ms training examples are generated from each cough by varying the location of the cough within each training example. This ensures the DNN is invariant to the position of a cough within the frame.

For each cough, two training examples are generated such



Fig. 3. Each 200 ms frame is subdivided into 50 ms windows - 42 MFCC features are calculated for each 50 ms segment.

that the beginning of the training example can occur anywhere within a 25 ms window before or after the cough start time (with uniform random probability). The remaining two out of four training examples are similarly generated, but the window is increased to 60 ms before or after the cough start.

Then, each 200 ms frame is further subdivided into four 50 ms windows to capture the temporal profile of each audio frame (Figure 3). From each 50 ms window we compute 42 features: 13 MFCCs, 13 MFCC delta features, and 13 MFCC delta-delta features. The remaining three features are the log energy within the 13 MFCCs, 13 MFCC delta features, and 13 MFCC delta-delta features. Since we break down each 200 ms frame into four 50 ms windows, we supply our network with 168 input features. To generate non-cough training examples, we randomly sample 200 ms segments of non-cough audio and calculate the same 168 features.

2.4. Neural Network Model for Cough Detection

Figure 4 shows our proposed neural network architecture. The DNN was trained with an equal number of positive and negative examples using stochastic gradient descent (SGD) and momentum. Using a grid search and cross-validation, we employed a learning rate of 0.15, momentum of 0.9, and a batch size of 150. The network was trained for 50 epochs.

3. RESULTS

Table 2 summarizes the leave-one-out specificity, sensitivity, and accuracy our algorithm achieves. The DNN is trained on all subjects except for one, which was left for testing. This is then repeated across all subjects in the entire dataset (leaveone-participant-out cross validation).

To compute accuracy, a test data set of four positive examples are created for each cough in accordance with the method

¹Less than 0.1% of all coughs were pruned



Fig. 4. Proposed network architecture for cough detection

outlined in Section 2.3. An equal number of negative test examples are randomly selected from background audio. To calculate sensitivity and specificity, a 200 ms sliding window is extracted from each 24 hour recording, with a step size of 50 ms. We define a false positive as any 200 ms frame that was incorrectly classified as a cough, and was not within 1 second of a cough.

The DNN achieves an average leave-one-out accuracy of 92.3%, with the highest accuracy of 96.2% for participant 5, and lowest accuracy of 89.7% for participant 7. This is due to the unique cough signature of participant 7 which is perceptually similar to throat clearing.

Sensitivity (true positive rate) and specificity (true negative rate) are more useful in describing the DNN's performance on 24 hour segments of audio due to the severely imbalanced classes. The algorithm results in an average specificity and sensitivity of 93.7% and 97.6% respectively.

Participant 4 and 7 both had the lowest specificity of 87%. This is due to the large amount of loud conversation in both of these recordings. Loud speech is one of the most likely sources of false positives. As several authors note, the most difficult part in designing ambulatory cough detection systems is robustness to false alarms [11] [12] [20]. Since classes

Subject	Specificity %	Sensitivity %	Accuracy %
1	97.7	92.2	95.2
2	95	97.4	95.4
3	88.1	97.8	91.9
4	87	97.3	91.2
5	97	97	96.2
6	98.3	97.6	94.3
7	87	94.4	89.7
8	96	96.1	92.5
9	97.3	97.6	92.3
Avg	93.7	97.6	92.3

 Table 2. Leave-one-out specificity, sensitivity, and accuracy of proposed algorithm



Fig. 5. Receiver Operating Characteristic (ROC) of our algorithm averaged across all participants. Area under the curve (AUC) is 0.93

are heavily imbalanced, specificity must be as high as possible to avoid large numbers of false positives.

As was expected, the sensitivity for participant 7 was lower than average (94.4%) due to the participant's uncommon cough pattern. The lowest leave-one-out sensitivity of 92.2% is found for participant 1. Since 3,133 out of the 5,670 (more than 55% of all training set coughs) come from participant 1, the DNN is likely to have an incomplete model of cough when the recording from participant 1 is left out from training.

The receiver operating characteristic (ROC) averaged across all participants is shown in Figure 5. We use the same test data set that is used to find accuracy in creating the ROC. The area under the curve (AUC) of the ROC is 0.93, a value close to 1 indicating that our model performs well in discriminating cough from background.

4. CONCLUSION

In this paper, we propose and implement a deep neural network and a data pre-processing algorithm for cough detection from ambulatory data collected with the FDA-cleared VitaloJAK device. We trained a DNN with two hidden layers on MFCC features to successfully discriminate coughing sounds from background noise. Results indicate our algorithm achieves high sensitivity, specificity, and accuracy on our extensive dataset. The proposed framework could decrease the load on medical personnel in labeling coughs from ambulatory audio recordings.

Future work will focus on extending the robustness of the algorithm with respect to the recording conditions and the recording device. Our goal is to use a similar algorithm to develop a reliable cough detector that can be used on audio data passively collected by personal mobile devices.

5. REFERENCES

- [1] S.M. Schappert, "National ambulatory medical care survey: summary," *Vital Health Statistics*, 1993.
- [2] P.G. Gibson, A.B. Chang, N.J. Glasgow, P.W Holmes, P. Katelaris, A.S. Kemp, L.I. Landau, S. Mazzone, P. New-combe, P. Van Asperen, and A.E. Vertigan, "CICADA: Cough in Children and Adults: Diagnosis and Assessment," *Medical Journal of Australia*, vol. 192, no. 5, pp. 265-71, 2010.
- [3] E. Young and J. Smith, "Quality of Life in Patients with Chronic Cough," *Therapeutic Advances in Respiratory Disease*, vol. 4, no. 5, pp. 49-55, 2010.
- [4] G. Comina, S. Larson, M. Bravard, et al., "Cough detection algorithm for monitoring patient recovery from pulmonary tuberculosis," *Proc. IEEE Engineering in Medical Biological Society Conf.*, pp. 6017-6020, 2011.
- [5] E. Kerem, S. Hirawat, S. Harmoni, Y. Yaakov, D. Shoseyov, M Cohen, et al. "Effectiveness of PTC124 treatment of cystic fibrosis caused by nonsense mutations: a prospective phase II trial," *Lancet*, vol. 372, pp. 719-727, 2008.
- [6] P.A. Marsden, J.A. Smith, A.A. Kelsall, E. Owen, J.R. Naylor, D. Webster, et al., "A comparison of objective and subjective measures of cough in asthma," *Journal of Allergy Clinical Immunology*, vol. 122, no. 5, pp. 903-907, 2008.
- [7] S.S. Birring, B. Prudon, A.J. Carr, et al., "Development of a symptom specific health status measure for patients with chronic cough: Leicester Cough Questionnaire (LCQ), "*Thorax*, vol. 58, pp. 339-343, 2003.
- [8] J.A. Smith and A. Woodcock, "New Developments in the Objective Assessment of Cough," *Lung*, vol. 186, pp. 48-54, 2007.
- [9] S.C. Decalmer, D. Webster, A.A. Kelsall, et al., "Chronic cough: how do cough reflex sensitivity and subjective assessments correlate with objective cough counts during ambulatory monitoring?" *Thorax*, vol. 62, pp. 329-334, 2007.
- [10] A. Barton, P. Gaydecki, K. Holt, and J.A. Smith, "Data Reduction for Cough Studies Using Distribution of Audio Frequency Content," *Cough*, vol. 8, no. 1, 2012.
- [11] S.S. Birring, T. Fleming, S. Matos, A.A. Raj, D.H. Evans, I.D. Pavord, "The Leicester Cough Monitor: preliminary validation of an automated cough detection system in chronic cough," *European Respiratory Journal*, vol. 31, pp. 1013-1018, 2008.

- [12] K. McGuinness, K. Holt, J.A. Smith, "Validation of the VitaloJAK 24-hour Ambulatory Cough Monitor," *Thorax*, 2012.
- [13] K. McGuinness, A. Morice, A. Woodcock and J. Smith, "The Leicester Cough Monitor: a semi-automated, semivalidated cough detection system?" *European Respiratory Journal*, vol. 32, no. 2, pp. 529-530, 2010
- [14] S. Barry, A. Dane, A. Morice, and A. Walmsley, "The automatic recognition and counting of cough," *Cough J.*, vol. 2, no. 8, 2006.
- [15] J. Amoh and K. Odame, "DeepCough: A Deep Convolutional Neural Network in a Wearable Cough Detection System," in *Proc. IEEE Biomedical Circuits and Systems Conf.*, 2015.
- [16] J. Amoh and K. Odame, "Deep Neural Networks for Identifying Cough Sounds, *IEEE Transactions on Biomedical Circuits and Systems*, vol. 10, 2015.
- [17] V. Swarnkar, U.R. Abeyratne, Y Amrulloh, C Hukins, R Triasih, A Setyati, "Neural Network Based Algorithm for Automatic Identification of Cough Sounds," *IEEE EMBC*, p. 1764-1767, 2013.
- [18] J.A. Smith et al., "The Description of Cough Sounds by Healthcare Professionals," *Cough*, vol. 2, no. 1, 2006.
- [19] J.M. Liu et al., "Cough Event Classification by Pretrained Deep Neural Network," *BMC Medical Informatics and Decision Making*, vol. 15, no. 4, 2015.
- [20] E.C Larson, T. Lee, S. Liu, M. Rosenfeld, and S.N. Patel, "Accurate and Privacy Preserving Cough Sensing using a Low-Cost Microphone," *Proc. of the 13th ACM International on Ubiquitous Computing Conf.*, pp. 375-384, 2011.