Can Machine Learning Be Used to Recognize and Diagnose Coughs?

Muhammad Nabeel
Universität Paderborn
14 PUBLICATIONS **38** CITATIONS

Charles John
University of Oklahoma
2 PUBLICATIONS **0** CITATIONS

Usama Masood
University of Oklahoma
8 PUBLICATIONS **5** CITATIONS

Haneya Qureshi
University of Oklahoma
8 PUBLICATIONS **4** CITATIONS

Can Machine Learning Be Used to Recognize and Diagnose Coughs?

Charles Bales[1], Muhammad Nabeel[1], Charles N. John[1], Usama Masood[1], Haneya N. Qureshi[1], Hasan Farooq[1], Iryna Posokhova[2,3], and Ali Imran[1,3]

[1] AI4Networks Research Center, University of Oklahoma, Tulsa, USA
[2] Kharkiv National Medical University, Kharkiv, Ukraine
[3] AI4Lyf LLC, USA
`csbales@wpi.edu`,
`{muhmd.nabeel,charlesj,usama.masood,haneya,hasan.farooq}@ou.edu`,
`iryna@ai4lyf.com`, `ali.imran@ou.edu`

Abstract. Emerging wireless technologies, such as 5G and beyond, are bringing new use cases to the forefront, one of the most prominent being machine learning empowered health care. One of the notable modern medical concerns that impose an immense worldwide health burden are respiratory infections. Since cough is an essential symptom of many respiratory infections, an automated system to screen for respiratory diseases based on raw cough data would have a multitude of beneficial research and medical applications. In literature, machine learning has already been successfully used to detect cough events in controlled environments. In this paper, we present a low complexity, automated recognition and diagnostic tool for screening respiratory infections that utilizes Convolutional Neural Networks (CNNs) to detect cough within environment audio and diagnose three potential illnesses (i.e., bronchitis, bronchiolitis and pertussis) based on their unique cough audio features. Both proposed detection and diagnosis models achieve an *accuracy* of over 89 %, while also remaining computationally efficient. Results show that the proposed system is successfully able to detect and separate cough events from background noise. Moreover, the proposed single diagnosis model is capable of distinguishing between different illnesses without the need of separate models.

Keywords: Convolutional neural network · cough detection · machine learning · mel-spectrogram · respiratory infection diagnosis.

1 Introduction

A study conducted in 2016 estimated that 4.4 % of all deaths in that year, a number surpassing 2 million, were due to various lower respiratory tract infections in both young children and adults [1]. Early detection of potential respiratory tract infections can reduce the likelihood of severe complications at a later time. Cough is an important and essential early symptom of many respiratory infections. Thus, an automated system for the detection and preliminary diagnosis of respiratory infections based on cough events would serve as a useful tool in the medical field.

Different types of audio categorization has been well investigated in current literature, to varying levels of success [2, 3]. Specifically, among the various machine learning based techniques, a Convolutional Neural Network (CNN) based approach has been a popular choice for this task [4, 5]. Using three publicly available datasets, ESC-50 [6], ESC-10 [6], and UrbanSound8K [7], that contain a large collection of environmental sounds, it is shown that a CNN model can be successfully applied to environmental sound classification [8]. This has been further verified by achieving similar results in [9].

Acknowledging the success of the CNN structure in categorizing audio into the fifty categories of the ESC-50 set, one of these being cough sounds, suggests that the network model may be capable of doing so with just coughs specifically. Indeed, this approach has been attempted in several other recent works [10–12]. In [10], authors utilized several feature types, such as Perceptual Linear Predictive (PLP) power spectrum and cepstrum, in their experiments to test the efficacy of the CNN structure in correctly identifying cough audio with image recognition. They used five different features to convert the sound files into image inputs and achieved an *accuracy* of 99.65 % suggesting that image processing stands as a valid method for categorizing cough audio. Later, authors in [11] achieved a *specificity* of 92.7 % using cough sounds processed into 64 ms segmented frames as inputs for their implemented CNN and deemed this method comparable to other conventional approaches. Finally, a recent work [12] used spectrogram images of lengthy audio segments as the inputs for a CNN to differentiate between healthy and pathological respiratory sounds, and achieved an *accuracy* of 86 %. Apart from CNNs, there are a plethora of potential approaches that can be utilized to tackle the issue, with the widely varying network model structures, cough sample sources and audio file processing [13, 14].

Each of these approaches involve either a heavy pre-processing phase or inherent design restrictions. These include but are not limited to discarding frames of silent audio through the use of an RMS energy threshold to only categorize selected audio [11], specifically selecting noise-reducing hardware for data collection [12] and carrying complex hardware [13]. Moreover, the categorization of audio inputs as different types of coughs requires a more complex machine learning model, making maintaining usefulness for mobile applications more difficult. This is because the minimization of processing power has been of less priority as compared to the model performance in these works. What this paper seeks to accomplish is to significantly minimize recorded audio pre-processing before feeding data into a network model while maintaining acceptably high levels of accuracy for cough detection, among other metrics. A primary difference in the inputs used here compared to other approaches is that this training data does not involve categorizing isolated cough events but rather deciding whether a select audio clip input contains one or more coughs within its length. One of the goals of this study is thus to first verify if a CNN model can recognize cough events by their distinct features among typical environmental noise, with minimum pre-processing requirements of cough samples. If a cough event is detected, the audio is then sent in for further processing to be used for the diagnosis portion of this research. The focus on low overhead for cough detection is for the sake of flexibility, allowing the detection model to potentially be run on devices with lower processing power than is typical, such as mobile devices.

For diagnosis of specific illnesses via cough audio authors in [15], exploited several machine learning models to accurately diagnose croup using only features extracted from cough audio, such as Mel-frequency cepstral coefficients. Similarly, authors in [16] proposed a low-complexity algorithm capable of identifying pertussis without any false diagnosis. They were able to achieve a high success rate through just the analysis of cough audio, their model recognizing features unique to the whooping cough illness. Results of similar other studies, such as [17, 18], suggest that coughs originating from specific infections or illnesses have a sufficient number of distinguishing features that machine learning models can use for categorization. Taking the next step of sorting through multiple potential causes simultaneously is what our work seeks to achieve.

The main contributions of this work can be summarized as follows:

- We propose a low-complexity CNN model for the detection of cough events from a typical audio segment with environmental noise, and then categorize among three common respiratory conditions with cough symptom.
- The dataset used to train our proposed model does not go through the typical level of pre-processing that other works use for cough detection.
- Contrary to prior works, e.g., [15, 16], that use a binary-class model to diagnose a particular disease (i.e., categorizing among that particular disease and normal class), our multi-class diagnosis model uses cough sound data to differentiate between several potential diagnoses that may be causing cough.

2 Cough Detection System

The effectiveness of a CNN typically scales with the size of the dataset used for training the model. Unfortunately, there is a particular scarcity of data in terms of cough audio. The results of prior works are difficult to compare with each other due to differing datasets used for model training, each with their own unique standardization. Cough data from a small sample of individuals [11, 19] will accurately identify coughs among these individuals and those similar, but may not be generally applicable to a larger populace. Studies with a large number of subjects and thus cough samples [12, 15] lead to accurate detection models, but datasets of this size for cough audio are typically collected in a controlled environment such as a hospital. Since the purpose of our CNN is to distinguish coughs within clips of environment sounds, controlled recording environments do not suit our purpose. Coughs cropped from YouTube audio [16] and other online sources, while numerous, vary wildly in audio quality. This variance, however, works towards the benefit of a more robust detection and diagnosis model that

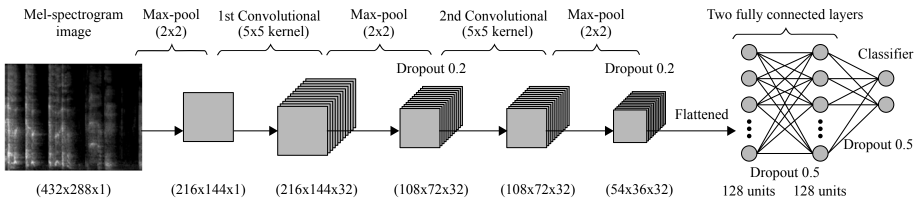

Fig. 1. Overview of the CNN structure used for cough detection.

is able to identify relevant cough audio in less controlled conditions. Our paper takes a similar approach to acquiring the data used for model training.

2.1 Detection Database Description

The database used for training the CNN in cough detection is composed of various modified audio clips gathered from open-source online sources [6, 20]. Each of these audio files originally contained at least one cough event and are cropped to a length of five seconds with the full cough contained at some arbitrary point within. Audio that contained coughs in more than a five seconds period are separated into multiple files. The exact number of coughs per file is left intentionally non-standardized. This database is balanced by a collection of environmental and speech sounds, comprised primarily of audio from the ESC-50 database for environmental sound classification. In addition to ESC-50, this non-cough half of the database is supplemented by sound clips taken from unused portions of the original cough audio that did not contain cough events. The final database contains 993 cough samples and 993 non-cough samples.

For our detection database, the input raw audio clips are transformed into Mel-spectrogram, resulting in a 2-dimensional image where one dimension represents time, other dimension represents frequency and the value of pixels in the image represent the amplitude. The resulting images of pixel size $432 \times 288 \times 3$ are then converted to gray-scale to unify the intensity scaling and are then compiled to form the final database of cough and no cough environmental audio clips.

2.2 Detection CNN Structure

The relative success of using a CNN for cough detection from recorded audio via image recognition serves as the basis for the machine learning model used for this system. Even though our goal is to analyze a relatively larger time frame of five seconds, the model structure serves as a workable foundation.

Overview of used CNN structure is shown in Fig. 1. Due to the high resolution of input image, we begin with a 2×2 max-pooling layer to lower input dimensions into the CNN. This will lower the required overall model complexity before proceeding. This is followed by a block of two convolutional layers, each having 32 filters of size 5×5, a stride of $(1, 1)$, ReLU activation function and a 2×2 max-pooling layer utilizing a 0.2 dropout. The features learned from this convolutional block are flattened before passing them to two fully connected layers, each having

128 neurons and ReLU activation function. These fully connected layers utilize a dropout of 0.5. The final layer is the softmax classification layer having 2 neurons to distinguish between cough and not cough for the given input.

The number of convolutional and fully connected layers are kept low to minimize potential overfitting issues. The database file count is increased to assist in reducing overfitting as well. Since ReLU is the current standard for CNNs, it is used for the activation functions of this model, while Adam [21] is used as the optimizer due to its relatively better efficiency and flexibility. A binary log loss function completes the detection model.

2.3 Detection Model Training and Results

The $1,986$ cough samples with 993 cough items and 993 non-cough items, are split into 70 % training, 15 % validation and 15 % testing datasets. The model is then trained with a batch size of 32 to find its optimal weights using an early stopping criteria on its validation dataset, so that the model training stops when it performs best on its validation dataset. Finally, the confusion matrix and the calculated performance metrics of the trained cough detection model on its testing dataset are reported in Table 1 and Table 2 respectively. The reported performance metrics are calculated as follows:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad , \tag{1}$$

$$specificity = \frac{TN}{TN + FP} \quad , \tag{2}$$

$$sensitivity/recall = \frac{TP}{TP + FN} \quad , \tag{3}$$

$$precision = \frac{TP}{TP + FP} \quad , \tag{4}$$

$$F1\text{-}score = 2 \times \left(\frac{precision \times recall}{precision + recall}\right) \quad . \tag{5}$$

Here, TP, TN, FP, and FN refer to True Positives, True Negatives, False Positives, and False Negatives, respectively. The FP is the value we wish to minimize the most to prevent any non-cough audio from being passed into the diagnosis model. Therefore, for this purpose, maximizing *specificity* is ideal.

These results reflect the performance of our system over a diverse dataset that comes from a comparatively huge swath of audio quality, cough volume and quantity. Much of the audio tested already contains ambient noise due to the varied range of sources used for collecting the data. Due to the large file count, potential biases such as environmental consistencies in instances where coughs would typically occur are lessened, and will continue to diminish as the dataset increases in size. Hence, it is important to highlight that the performance of our proposed detection system is likely to improve further, for example, in controlled environments, which were typically the focus of prior works.

		Predicted	
		No Cough	Cough
Actual	No Cough	86.2	13.8
	Cough	8.1	91.9

Table 1. Normalized confusion matrix for cough detection (in percentage).

F1-Score (%)	Sensitivity (%)	Specificity (%)	Precision (%)	Accuracy (%)
89.35	91.9	86.2	86.94	89.05

Table 2. Performance Metrics for Cough Detection.

3 Cough Diagnosis System

The data scarcity issue, as in the case of cough detection dataset, is also present for the diagnosis model, but compounded with the additional problem of labeled cough audio being considerably more scarce. Since the existing database for detection is comprised primarily of unlabeled cough audio with no identified illness relation, the same cannot be used as diagnosis training data. To train our cough diagnosis system, we collected cough sounds from 35 bronchiolitis, 131 pertussis and 102 bronchitis patients. The performance and reliability of the diagnosis system is likely to improve as more data becomes available. However, even with small training data, very promising results have been observed on unseen test samples, as reported later in diagnosis results section.

3.1 Diagnosis Database Composition

Since the purpose of the diagnosis model is not to detect the presence of any number of coughs but to categorize cough events according to their subtle differences, the input audio files for this model are much shorter in length. Hence, we adopt a different data preparation process for making it suitable for training.

The sound files are first cut to a single cough event with the longest of these files lasting two seconds, the rest are buffered to the same length with noiseless audio to maintain a fixed two second length for all samples. This fixed length is set for the spectrogram conversion process that comes afterwards and all the two-second audio files are converted into $432 \times 288 \times 1$ Mel-spectrograms images. Finally, all images in the cough diagnosis dataset are then converted to gray-scale to unify its intensity scaling and reduce the size of each sample to facilitate the model training process.

3.2 Diagnosis CNN Structure

The classification model used for cough detection can also be applied to the labeled cough data originating from several different illnesses. However, due to the subtle differences between two coughs of differing illnesses, as compared to cough and non-cough event, a more complex set of layering is required for this cough diagnosis model.

The structure of the CNN used for diagnosis is similar to the detection CNN structure in the beginning (depicted in Fig. 1). However, after the 2×2 max-pooling layer at the end of convolutional block, another similar convolutional block comprising of two convolutional layers is added before being passed on to a similar structure of two fully-connected layers with an identical neuron count of 128, activation type, and dropout of 0.5. However, the final output layer now has 3 neurons and Softmax activation function to classify the input between three possible diseases.

3.3 Diagnosis Model Training and Results

The diagnosis database is comprised of 268 total cough sound items, distributed among bronchiolitis, pertussis, and bronchitis at counts of 35, 131, and 102, respectively. The database is again split into 70 % training, 15 % validation and 15 % testing sets. The model is trained using an early stopping criteria on its validation dataset, so that the model training stops when it performs best on its validation dataset. Finally, the confusion matrix and the calculated performance metrics of the trained cough detection model on its testing dataset are reported in Table 3 and Table 4 respectively.

This performance metrics clearly shows that the ability of this machine learning model to distinguish between multiple illnesses based on various cough sounds. It is also important to note that in previous works, the focus is usually to diagnose a specific illness from cough sounds, whereas we distinguish between multiple types of disease coughs in the same model.

Due to the unbalanced nature of the database used for training the diagnosis model, *F1-Score* is more important here rather than the *accuracy* for making conclusions on the model performance. Since the data count is relatively low for the number of categories, more than one metric better interprets model performance. The potential bias mentioned for detection, environmental consistencies between disease categories unrelated to the cough audio structure, is also a relevant concern in this case. For example, individuals with a known case of bronchitis would typically be situated in hospitals more frequently than those with the flu or a cold. This issue would again be alleviated by a larger volume of data, or perhaps a controlled variance in making sure that one category of data is not overtly reliant on background sounds for its unique model training identifiers. Regardless, the achieved performance, given the number of disease categories, is indicative of high level of success in diagnosing between specific illness-based coughs.

		Predicted		
		Bronchiolitis	Bronchitis	Pertussis
Actual	Bronchiolitis	80.0	20.0	0.0
	Bronchitis	0.0	93.8	6.2
	Pertussis	0.0	5.0	95.0

Table 3. Normalized confusion matrix for cough detection (in percentage).

	F1-Score (%)	**Sensitivity** (%)	**Specificity** (%)	**Precision** (%)	**Accuracy** (%)
Overall	-	-	-	-	89.60
Pertussis	94.43	95.00	96.90	93.87	-
Bronchitis	85.74	93.80	87.50	78.95	-
Bronchiolitis	88.89	80.00	100.00	100.00	-

Table 4. Performance Metrics for Cough Diagnosis Classifier.

4 Conclusion

We present a low complexity, automated recognition and diagnostic tool for screening respiratory infections that utilizes a CNN-based approach to detect cough within various types of environment noise and diagnose three potential illnesses (i.e., bronchitis, bronchiolitis and pertussis) based on their unique cough audio features. The approach uses cough audio and converts it to Mel-spectrograms for both detection and diagnosis.

Experiments show that CNN is capable of accomplishing these tasks with a high level of *accuracy*. With only a small number of modifications, the low-complexity network model can be trained sufficiently for both ends. This is achieved in spite of limited available cough data. The model performance is likely to improve given a larger training dataset.

In future work, the focus is to implement the detection model on mobile devices for a more consistent data acquisition method via phone recording. Additionally, increasing the dataset size for diagnosis as well as affirmation of manual labeling *accuracy* through work with an experienced physician will both improve the reliability of diagnosis model and scale it for diagnosing a wider range of potential cough-based illnesses.

References

1. C. Troeger, B. Blacker, I. A. Khalil, P. C. Rao, J. Cao, S. R. M. Zimsen, S. B. Albertson, A. Deshpande, T. Farag, Z. Abebe *et al.*, "Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory

infections in 195 countries, 1990–2016: a systematic analysis for the global burden of disease study 2016," *The Lancet Infectious Diseases*, vol. 18, no. 11, pp. 1191–1210, Nov. 2018.

2. B. Ghoraani and S. Krishnan, "Time–frequency matrix feature extraction and classification of environmental audio signals," *IEEE transactions on audio, speech, and language processing*, vol. 19, no. 7, pp. 2197–2209, 2011.
3. J. R. Delgado-Contreras, J. P. García-Vázquez, and R. F. Brena, "Classification of environmental audio signals using statistical time and frequency features," in *2014 International Conference on Electronics, Communications and Computers (CONIELECOMP)*. IEEE, 2014, pp. 212–216.
4. S. Hershey, S. Chaudhuri, D. P. W. Ellis, J. F. Gemmeke, A. Jansen, R. C. Moore, M. Plakal, D. Platt, R. A. Saurous, B. Seybold, M. Slaney, R. J. Weiss, and K. Wilson, "CNN Architectures for Large-scale Audio Classification," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. New Orleans, LA: IEEE, Mar. 2017, pp. 131–135.
5. T. Kim, J. Lee, and J. Nam, "Comparison and Analysis of SampleCNN Architectures for Audio Classification," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 2, pp. 285–297, May 2019.
6. K. J. Piczak, "ESC: Dataset for Environmental Sound Classification," in *23rd ACM International Conference on Multimedia*, ser. MM 15. Brisbane, Australia: ACM, Oct. 2015, pp. 1015–1018.
7. J. Salamon, C. Jacoby, and J. P. Bello, "A dataset and taxonomy for urban sound research," in *Proceedings of the 22nd ACM international conference on Multimedia*, 2014, pp. 1041–1044.
8. K. J. Piczak, "Environmental sound classification with convolutional neural networks," in *2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, 2015, pp. 1–6.
9. Z. Zhang, S. Xu, S. Cao, and S. Zhang, "Deep Convolutional Neural Network with Mixup for Environmental Sound Classification," in *Pattern Recognition and Computer Vision*, J.-H. Lai, C.-L. Liu, X. Chen, J. Zhou, T. Tan, N. Zheng, and H. Zha, Eds. Springer International Publishing, 2018, pp. 356–367.
10. H.-H. Wang, J.-M. Liu, M. You, and G.-Z. Li, "Audio Signals Encoding for Cough Classification Using Convolutional Neural Networks: A Comparative Study," in *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. Washington, D.C.: IEEE, Nov. 2015, pp. 442–445.
11. J. Amoh and K. Odame, "Deep Neural Networks for Identifying Cough Sounds," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 10, no. 5, pp. 1003–1011, Oct. 2016.
12. M. Aykanat, Ö. Kiliç, B. Kurt, and S. Saryal, "Classification of Lung Sounds Using Convolutional Neural Networks," *EURASIP Journal on Image and Video Processing*, vol. 2017, no. 1, p. 65, Sep. 2017.
13. M. Młyńczak, K. Pariaszewska, and G. Cybulski, "Automatic cough episode detection using a vibroacoustic sensor," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2015, pp. 2808–2811.
14. M. You, H. Wang, Z. Liu, C. Chen, J. Liu, X.-H. Xu, and Z.-M. Qiu, "Novel feature extraction method for cough detection using nmf," *IET Signal Processing*, vol. 11, no. 5, pp. 515–520, 2017.
15. R. V. Sharan, U. R. Abeyratne, V. R. Swarnkar, and P. Porter, "Automatic Croup Diagnosis Using Cough Sound Recognition," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 2, pp. 485–495, Feb. 2019.

16. R. X. A. Pramono, S. A. Imtiaz, and E. Rodriguez-Villegas, "A Cough-Based Algorithm for Automatic Diagnosis of Pertussis," *PLOS ONE*, vol. 11, no. 9, p. e0162128, Sep. 2016.
17. C. Infante, D. Chamberlain, R. Fletcher, Y. Thorat, and R. Kodgule, "Use of cough sounds for diagnosis and screening of pulmonary disease," in *2017 IEEE Global Humanitarian Technology Conference (GHTC)*. IEEE, 2017, pp. 1–10.
18. P. Porter, U. Abeyratne, V. Swarnkar, J. Tan, T.-w. Ng, J. M. Brisbane, D. Speldewinde, J. Choveaux, R. Sharan, K. Kosasih *et al.*, "A prospective multicentre study testing the diagnostic accuracy of an automated cough sound centred analytic system for the identification of common respiratory disorders in children," *Respiratory research*, vol. 20, no. 1, p. 81, 2019.
19. J. Liu, M. You, G. Li, Z. Wang, X. Xu, Z. Qiu, W. Xie, C. An, and S. Chen, "Cough Signal Recognition with Gammatone Cepstral Coefficients," in *IEEE China Summit and International Conference on Signal and Information Processing*. Beijing, China: IEEE, July 2013, pp. 160–164.
20. F. Font, G. Roma, and X. Serra, "Freesound Technical Demo," in *21st ACM International Conference on Multimedia*, ser. MM 13. Barcelona, Spain: ACM, Oct. 2013, pp. 411–412.
21. D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *International Conference on Learning Representations (ICLR)*, San Diego, CA, May 2015.