깊은 시계열 특성 추출을 이용한 폐 음성 이상 탐지

Kim-Ngoc T. Le[1], 변규린 [1], 추현승 [1,2]
[1] 성균관대학교 AI 시스템공학과
[2] 성균관대학교 전자전기컴퓨터공학과
ltkngoc0228@skku.edu, byungyurin21@g.skku.edu, choo@skku.edu

Detection of Anomaly Lung Sound using Deep Temporal Feature Extraction

Kim-Ngoc T. Le[1], Gyurin Byun[1], Hyunseung Choo[1,2]
[1]Dept. of AI Systems Engineering, Sungkyunkwan University
[2]Dept. of Electrical and Computer Engineering, Sungkyunkwan University

Abstract

Recent research has highlighted the effectiveness of Deep Learning (DL) techniques in automating the detection of lung sound anomalies. However, the available lung sound datasets often suffer from limitations in both size and balance, prompting DL methods to employ data preprocessing such as augmentation and transfer learning techniques. These strategies, while valuable, contribute to the increased complexity of DL models and necessitate substantial training memory. In this study, we proposed a streamlined and lightweight DL method but effectively detects lung sound anomalies from small and imbalanced dataset. The utilization of 1D dilated convolutional neural networks enhances sensitivity to lung sound anomalies by efficiently capturing deep temporal features and small variations. We conducted a comprehensive evaluation of the ICBHI dataset and achieved a notable improvement over state-of-the-art results, increasing the average score of sensitivity and specificity metrics by 2.7%.

1. Introduction

Automation of analysis aims to mitigate the variability associated with human interpretation, providing a more consistent and objective diagnosis [1]. Recently, deep learning methods have gained prominence in this field. Deep Neural Networks (DNNs), in particular, have demonstrated their potential in automating the analysis of respiratory sounds [1-3]. DNNs excel at learning intricate patterns in high-dimensional data, making them well-suited for identifying anomalies in the inherently complex and non-linear domain of respiratory sounds. Several studies have explored state-of-the-art architectures, including ResNet [2-4], and CNN-MoE [3], to achieve effective anomaly detection in respiratory sound.

A central concern pertains to the substantial data requirements of Deep Neural Networks (DNNs) for effective learning. Consider the largest publicly available dataset for respiratory sounds, the International Conference on Biomedical Health Informatics (ICBHI) dataset, which comprises a mere 6898 breathing cycles [5]. Given the intricate nature of respiratory sounds and their representation of a wide range of pathological conditions, this dataset falls notably short in facilitating the training of robust DNN models. Furthermore, the ICBHI dataset suffers from imbalances, prompting the application of data preprocessing techniques such as augmentation and transfer learning within the domain of Deep Learning (DL) [1-3]. Although these methods are invaluable, they introduce heightened complexity to DL models and necessitate substantial training memory. As a result, it becomes quite challenging to incorporate the deep learning frameworks in the currently available wearable devices and mobile platforms.

In light of these challenges, we introduced a streamlined and lightweight DL method but effectively identifies lung sound anomalies from small and imbalanced dataset. Leveraging 1D dilated convolutional neural networks enhances sensitivity to these anomalies by efficiently capturing extended temporal patterns and subtle variations. Our comprehensive evaluation on the ICBHI dataset resulted in a notable improvement over state-of-the-art results, with a 2.7% increase in the average score of sensitivity and specificity metrics. Notably, our method achieved an impressive accuracy of 67.4%, surpassing other state-of-the-art models. Furthermore, we conducted experiments with different sample lengths and found that 15000 samples performed the best. Smaller sample lengths caused sample clipping, leading to the loss of valuable information in an already limited dataset, while larger sample lengths resulted in repetition, decreasing performance.

2. The proposed method

In this section, we prensent a lightweight and simple model for respiratory anomaly detection. The architectural layout of our proposed approach is depicted in Fig. 1, consisting of two primary components: the front-end feature extraction and the back-end neural network classifier. The workflow initiates by transforming respiratory cycles extracted from lung sound signals into sequence-based representations. These sequences then serve as input to the scheme, which categorizes them as either normal or exhibiting abnormal patterns.

As shown in Figure 1, the network architecture consists of seven blocks. Six of these blocks are composed of 1D dilated convolutional layers with varying parameters: kernel size of 2 x 2 and kernel numbers of {64, 128, 256, 256, 512, 512}, each with a dilation rate of 2. These blocks also incorporate rectified linear units (ReLU), average pooling (AP) with a kernel size of 2 x 2, and global average pooling (GAP) layers. The seventh block is a dense block, featuring a fully connected (FC) layer and a final Softmax layer for classification.

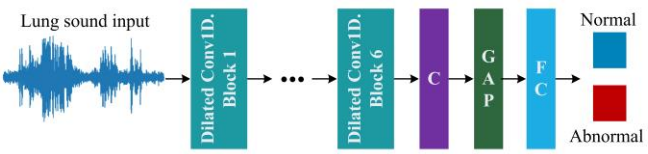

Fig 1. The proposed method.

3. Experiment setting and Performance evaluation.

3.1 Dataset and Evaluation metrics

Dataset: Our evaluations utilize the ICBHI Scientific Challenge Respiratory Sound Dataset, comprising 5.5 hours of recordings. These respiratory recording categories are distributed across 920 audio samples obtained from 126 subjects. The ICBHI dataset encompasses a total of 6898 respiratory cycles with the number of cycles being unbalanced, including 1864, 886, 506 and 3642 cycles respectively for *crackle*, *wheeze*, *crackle and wheeze*, and *normal*. The length of breathing cycles ranges from 0.2s to 16.2s. This poses a problem while training our network as it expects a fixed size input. Before feature extraction, we initiate the process by resampling the cycle recordings to a 4kHz frequency. To ensure uniformity in input data for our deep learning models, we employ a standard technique known as zero-padding. This approach is utilized to standardize the length of the sample sequence data to a predetermined size. When the sequence is shorter than desired, we pad it by adding zeros to reach the desired length. Conversely, if the sequence is equal to or longer than desired, we truncate it to match the desired length.

Within this study, we conduct experiments using various sample sequence lengths to determine the optimal input size.

Metrics: Following the official ICBHI 2017 challenge guidelines [5], we adopt the same evaluation metrics. The evaluation score, denoted as S_c, is determined as the mean of the sensitivity S_e and specificit S_p scores. This approach provides a comprehensive assessment of our model's performance by taking into account both true positive and true negative rates. The S_e and S_p scores are computed using the following formulas:

$$S_e = \frac{P_c + P_w + P_b}{T_c + T_w + T_b}, \qquad S_p = \frac{P_n}{T_n}, \qquad S_c = \frac{S_e + S_p}{2}$$

where P_c, P_w, P_b represent the numbers of correct predictions for the crackle, wheeze, normal, and both classes, respectively, while T_c, T_w, T_b, T_n correspond to the total number of instances of each class.

3.2 Experiment results

This study uses a random 60/40 train/test splitting strategy recommended by the ICBHI challenge. We conduct a comparative analysis between our proposed framework and state-of-the-art systems such as ResNet-Att [1], CNN-MoE [2] and ResNet50 [3] to gauge its efficacy in anomaly sound classification. Through the examination of results obtained from a train/test splitting approach, our objective is to glean insights into the influence of the splitting strategy on the overall performance of the classification model.

Table 1. Performamce comparison between the proposed scheme and state-of-the-art schemes following the random data split (Highest scores are highlighted in bold)

Methods	S_p	S_e	S_c	Parameter
ResNet-Att [1]	71.4	51.4	61.4	25M
CNN-MoE [2]	72.4	37.5	54.1	14M
ResNet50 [3]	**79.3**	50.1	64.7	23M
Our	67.2	**67.6**	**67.4**	**1M**

In the domain of respiratory anomaly detection, our proposed method demonstrates superior performance compared to other systems, as evident from the overall ICBHI score. However, it may not consistently outperform them simultaneously for both subcomponents, namely specificity and sensitivity. As highlighted in Table 1, our method exhibits a substantial improvement over state-of-the-art approaches, with an increase of over 2.7%. Notably, our model achieves an impressive accuracy of 67.4%, surpassing other state-of-the-art models.

Deep Learning (DL) models find extensive applications in real-world scenarios, but they do come with certain drawbacks, such as a high number of parameters that demand substantial computational resources. Additionally, DL models typically require longer training times compared to machine learning (ML) models, often 2 to 4 times longer. Interestingly, our experiments reveal that increasing the number of parameters in DL models does not necessarily lead to improved performance. For example, even though ResNet101 has twice the number of parameters compared to ResNet50, its performance remains lower. This suggests that a deep and dense neural network is not essential for our respiratory anomaly detection task. Our model achieves high accuracy with significantly fewer parameters, specifically 1M parameters for training, resulting in reduced training time.

Utilizing 1D dilated convolutional layers with tailored parameters, such as a kernel size of 2 x 2 and a dilation rate of 2, brings significant advantages to the task of anomaly detection in lung sound audio. The dilation rate of 2 endows these layers with the capability to capture extended temporal patterns within the audio data, a crucial requirement for detecting subtle respiratory anomalies. In the realm of lung sound analysis, where pathological conditions may manifest over varying time scales, this ability to consider broader contexts proves invaluable. Furthermore, the selected parameters present a balance between model complexity and computational efficiency, making real-time analysis feasible, even when dealing with substantial audio datasets. This approach effectively extracts relevant features, providing the foundation for precise and robust anomaly detection, ultimately advancing the diagnosis and monitoring of respiratory conditions with greater accuracy and efficiency.

Moreover, we conducted experiments with varying sample lengths and identified that 15000 samples yielded the best results, shown as Table 2. Smaller sample lengths led to sample clipping, resulting in the loss of valuable information within our already limited dataset. Larger cycle lengths resulted in repetition, which adversely affected performance.

Table 2. Lung sound sample length size versus classification score
(Highest scores are highlighted in bold)

Length S_C	5000 samples	10000 samples	15000 samples	20000 samples
	57.0	61.2	**67.4**	62.8

4. Conclusion

This paper introduces a robust and lightweight framework designed for accurate respiratory anomaly detection from lung auscultation recordings, even when confronted with limited dataset sizes. Leveraging innovative feature exploitation techniques, our proposed method surpasses the state-of-the-art, showcasing improvements of over 2.7% in terms of Sc for the 60/40 train/test splitting. Additionally, our investigations into different sample lengths highlight that 15000 samples yielded optimal results. However, our respiratory anomaly detection result is still low with 67.4% of Sc, so applying the curren proposal in real-world scenarios remains challenging. In future work, we will consider improving our proposed framework by using 2D dilated convolution to detect potential abnormalities in these signals.

References

[1] Himadri Mukherjee, Priyanka Sreerama, Ankita Dhar, Sk Md Obaidullah, Kaushik Roy, Mufti Mahmud, and KC Santosh. Automatic lung health screening using respiratory sounds. Journal of Medical Systems, 45:1–9, 2021.

[2] Lam Pham, Huy Phan, Ramaswamy Palaniappan, Alfred Mertins, and Ian McLoughlin. Cnn-moe based framework for classification of respiratory anomalies and lung disease detection. IEEE journal of biomedical and health informatics, 25(8):2938–2947, 2021.

[3] Truc Nguyen and Franz Pernkopf. Lung sound classification using co-tuning and stochastic normalization. IEEE Transactions on Biomedical Engineering, 69(9):2872–2882, 2022.

[4] Siddhartha Gairola, Francis Tom, Nipun Kwatra, and Mohit Jain. Respirenet: A deep neural network for accurately detecting abnormal lung sounds in limited data setting. In 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pages 527–530. IEEE, 2021.

[5] BM Rocha, Dimitris Filos, L Mendes, Ioannis Vogiatzis, Eleni Perantoni, E Kaimakamis, P Natsiavas, Ana Oliveira, C J´acome, A Marques, et al. A respiratory sound database for the development of automated classification. In Precision Medicine Powered by pHealth and Connected Health: ICBHI 2017, Thessaloniki, Greece, 18-21 November 2017, pages 33–37. Springer, 2018.

Acknowledgement

This work is supported by an IITP grant funded by the Korean government (MSIT) under the ICT Creative Consilience program (IITP-2023-2020-0-01821), Artificial Intelligence Innovation Hub (IITP-2021-0-02068), and AI Graduate School program (IITP-2019-0-00421).