Pathological Lung Sound Detection using Deep Transfer Learning

IJIMSR, Vol. 2, No. 1, (2024) 24.2.1.010

Smitha Raveendran
Ramrao Adik Institute of Technology
D Y Patil Vidyanagar, Sector 7, Nerul, Navi Mumbai
Email : smitha.raveendran@rait.ac.in

Jitendra Sonawane
Ramrao Adik Institute of Technology
D Y Patil Vidyanagar, Sector 7, Nerul, Navi Mumbai
Email : jitendra.sonawane@rait.ac.in

Gajanan K. Birajdar
Ramrao Adik Institute of Technology
D Y Patil Vidyanagar, Sector 7, Nerul, Navi Mumbai
Email: gajanan.birajdar@rait.ac.in

***Mukesh D. Patil**
Department of Electronics and Telecommunication Engineering
Ramrao Adik Institute of Technology, DY Patil Deemed to be University
Nerul, Maharashtra, 400706, India
Email: mukesh.patil@rait.ac.in

***Corresponding author**
Received 2nd September 2023; Accepted 14th Mar 2024
KEYWORDS: Deep Learning, Lung Sound, Spectrogram, Alexnet, Googlenet, Performance Metrics

Pathological Lung Sound Detection using Deep Transfer Learning

Smitha Raveendran
Ramrao Adik Institute of Technology
D Y Patil Vidyanagar, Sector 7, Nerul, Navi Mumbai
Email : smitha.raveendran@rait.ac.in

Jitendra Sonawane
Ramrao Adik Institute of Technology
D Y Patil Vidyanagar, Sector 7, Nerul, Navi Mumbai
Email : jitendra.sonawane@rait.ac.in

Gajanan K. Birajdar
Ramrao Adik Institute of Technology
D Y Patil Vidyanagar, Sector 7, Nerul, Navi Mumbai
Email: gajanan.birajdar@rait.ac.in

Mukesh D. Patil
Department of Electronics and Telecommunication Engineering
Ramrao Adik Institute of Technology, DY Patil Deemed to be University
Nerul, Maharashtra, 400706, India
Email: mukesh.patil@rait.ac.in

ABSTRACT

Lung sound analysis has gained prominence as a non-invasive method for diagnosing respiratory conditions. Recent development in deep transfer learning models have signified the potential to enhance the accuracy of lung sound detection, enabling early and accurate diagnosis. This paper presents an approach for lung sound detection using deep transfer learning techniques. A deep neural network architecture pretrained on a large external dataset and fine-tuned on a specialized lung sound dataset to leverage both general and domain-specific features. Firstly, input lung sound recordings are transformed into three spectrogram images. Two transfer learning models AlexNet and GoogleNet captures intricate patterns within lung sounds, differentiating between normal respiratory sounds and those indicative of pathological conditions. To evaluate the effectiveness of our approach, several comprehensive experiments are performed on a ICBHI 2017 dataset of lung sounds. The results showcase the improved performance of deep transfer learning model compared to conventional methods and standalone deep learning architectures with a significant reduction in false positives and false negatives. Highest detection rate of 94.64% is attained by GoogleNet model on ICBHI database.

KEYWORDS: Deep Learning, Lung Sound, Spectrogram, Alexnet, Googlenet, Performance Metrics

1. INTRODUCTION

According to Li et al. (2020), respiratory disease that results from lung anomalies accounts for 7% of global mortality. Chronic respiratory conditions (CRC) affect the lungs, their supporting tissues, and the airways.Asthma and chronic obstructive pulmonary disease (COPD) are the most prevalent CRDs .Tobacco smoking, occupational dust, extreme air pollution, presence of chemicals in atmosphere all contribute to various respiratory tract diseases. Routine patient examinations aid in prompt detection of CRDs because they are not entirely treatable. The patient is then given a suitable treatment regimen, which can assist manage their symptoms and improve their quality of life [1].

Lung auscultation continues to be a crucial component of the physical examination since it is an inexpensive, non-invasive, and safe diagnostic method. It is still highly susceptible to erroneous interpretations as it relies on the health care professional and the reliability of diagnostic instruments. Traditional approaches rely on issues like qualified and well-trained healthcare professionals and the cooperation of the patients. Timely detection of lung disorders are extremely important as delay in treatment may even lead to death as it relies on the health care professional and the reliability of diagnostic instruments.

A useful source of data for the investigation of pulmonary pathology is lung sounds. Lung sounds are difficult to analyse and discriminate because they are non-stationary and non-linear signals [2]. Doctors investigate the features of lung sound using a variety of approaches. The tool most frequently used to diagnose pathological problems in human lungs is the stethoscope. Although X-rays are a less expensivealternative, the radiation they emit is damaging to the body. Having an automatedsystem that can effectively analyse pathological situations utilising lung sounds is therefore advised.

Although early research emphasized on manually created features and conventional machine learning, more recent works have focused more on deep learning-based methods [3]. In this work an approach for lung sound detection using deep learning architectures is implemented, that will enable us to quickly and accurately identify lung abnormalities. The method employed in this work utilizes deep learning architectures for lung sound classification.

Leveraging the power of deep learning, this innovative technique aims to identify and classify abnormal lung sounds, such as wheezing or crackling, from audio recordings with better accuracy. By employing transfer learning, the model can leverage pre-trained neural networks on extensive datasets to adapt and excel in the specific task of lung sound analysis. This technology holds immense promise for early disease detection, enabling healthcare professionals to swiftly diagnose respiratory conditions, thereby improving patient outcomes and revolutionizing the field of pulmonary medicine.

In this work, the respiratory sound recordings from ICBHI 2017 challenge dataset is used for all evaluations. The collection consists of 920 recordings from 126 patients totaling 5.5 hours in length [4]. The audio recordings are converted to spectrogram images. Conventional spectrogram has been used in this work. Spectrogram time-frequency images obtained from audio samples of lung sounds are used for classification. Pre-trained networks from transfer learning architectures are used for feature extraction and classification. The deep transfer learning models are used for the classification of lung sounds into normal and abnormal samples.

The work is organized as follows: Section II lists the literature and background used in the work. Section III describes the proposed algorithm. Section IV presents the deep learning models and in section V, Experimental results are presented. Section VI concludes the work.

2. LITERATURE SURVEY

In [5], a novel framework for respiratory sound classification through the synergistic integration of Empirical Mode Decomposition (EMD) and Bidirectional Long Short-Term Memory (BiLSTM) networks is investigated. The proposed methodology addresses the intricate challenge of analyzing respiratory sounds by leveraging the intrinsic characteristics of the sound data.

The article [6] critically examines the advancements and implications of employing deep learning models in the context of detecting respiratory pathologies from raw lung auscultation sounds. A systematic analysis of the paper reveals the pivotal role played by Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and attention mechanisms in enhancing the diagnostic accuracy of respiratory disorders.

The study presented in [7] amalgamates a combination of SVM, LSTM, and Bayesian optimization algorithms, accentuating the potential synergy between traditional and deep learning methodologies. The method is underpinned by an intricate feature extraction technique leveraging wavelet bi-phase and bi-spectrum, showcasing multi-dimensional aspects of respiratory sounds.

[8]proposed "ALSD-Net," an innovative automatic lung sounds diagnosis network from pulmonary signals. The study elucidates the intricate architecture designed to revolutionize the realm of respiratory sound analysis. The "ALSD-Net" introduces a deep learning model

strategically engineered to classify and diagnose pulmonary conditions accurately.

Authors in [9] presented a detailed exploration of the innovative "CNN-MoE" framework, designed for classification of respiratory anomalies and lung disease detection. The study delves into the novel approach of combining Convolutional Neural Networks (CNNs) with a Mixture of Experts (MoE) architecture to enhance the accuracy and efficiency of respiratory sound analysis. The algorithmic underpinning of the "CNN-MoE" framework is meticulously dissected, revealing the synergy between CNNs' feature extraction capabilities and MoE's adaptability to different subtasks. A new approach involves the application of Convolutional Neural Networks (CNNs) and employs snapshot ensembling techniques to enhance classification accuracy [10]. The review dissects the underlying architecture, elucidating the sequential implementation of snapshot ensembling to boost model performance. Pre-processing steps, encompassing spectrogram generation and augmentation strategies, are expounded, highlighting their role in feature extraction and dataset augmentation.

The work in [11] developed an approach based on a non-local block ResNet Neural Network coupled with mix-up data augmentation to enhance the classification of adventitious lung sounds. The empirical evaluation presented in the research validates the efficacy of the "LungRN+NL" approach, showcasing marked improvement in the classification of lung pathologies. An innovative approach in [12] that amalgamates machine learning techniques with wavelet transform for precise classification of respiratory signals. The work systematically analyses the algorithmic intricacies of the proposed framework, emphasizing the strategic integration of wavelet transform to extract multi-scale features from the respiratory signals. Furthermore, the study delves into the machine learning models deployed, encompassing Support Vector Machines, Random Forests, and neural networks, all of which capitalize on the feature-rich wavelet coefficients.

3. PROPOSED ALGORITHM

A lung sound classification system is modeled using time-frequency visual representations. Conventional spectrogram is used for the visual presentation of the abnormality present in the sound samples. The detailed block diagram of the lung sound classification is shown in figure 3.1.

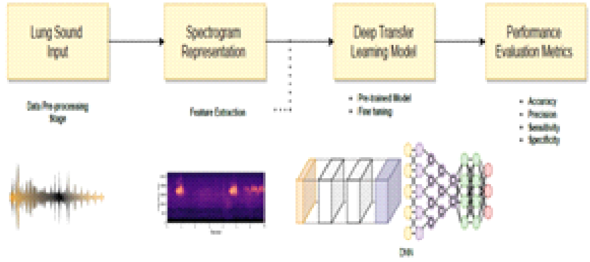

Figure 3.1. Framework for Pathological Lung Sound Detection using Deep Learning Framework

The lung sound signals after pre-processing are used to generate spectrogram images. These images are given as input to deep transfer learning models for classification. The performance of the models is assessed by computing the different evaluation metrics.

4. DEEP LEARNING

Deep learning has emerged as a transformative technology in the field of biomedical signal processing, offering improved capabilities for the analysis and interpretation of complex physiological data. Biomedical signal processing encompasses a wide range of applications, from electrocardiograms (ECGs) and electroencephalograms (EEGs) to medical imaging and genomic data. Deep learning algorithms, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have shown remarkable proficiency in handling these signals.

In the domain of medical imaging, CNNs have revealedoutstanding performance in tasks such as medical image segmentation, object detection, and disease classification. They can automatically extract relevant features from radiological images like X-rays, CT scans, and MRIs, aiding radiologists in faster and more accurate diagnosis. They can effectively detect anomalies, predict patient outcomes, and assist in real-time monitoring of critical conditions like arrhythmias and seizures [3].

CNN models used for lung sound classification shows better performance thanconventional methods of classification. For deep-feature extraction, a pretrained CNN model that takes spectrogram images as input is used in this work [13].

AlexNet and GoogleNet are both influential convolutional neural network (CNN) architectures that have made significant contributions to the field of deep learning and computer vision are employed in this work. Eight layers make up the AlexNet architecture: five convolutional layers and three fully connected layers.A 1000-way softmax receives the output of the final fully connected layer and generates a distribution across the 1000 class labels.

GoogleNet, also known as the Inception architecture, was developed by researchers at Google and won the ILSVRC in 2014. Its primary innovation lies in the use of inception modules, which allow for the efficient use of computation resources while maintaining high accuracy. Key features of GoogleNet include:

4.1 Inception Modules: GoogleNet introduced the concept of inception modules, which are comprised of multiple parallel convolutional layers of different filter sizes and depths. This architecture enables the network to capture features at multiple scales without significantly increasing computational complexity.

4.2 1x1 Convolutions: The use of 1x1 convolutions within the inception modules helps reduce the dimensionality of the feature maps, thus saving computational resources.

4.3 Global Average Pooling: Instead of fully connected layers at the end, GoogleNet employs global average pooling, reducing the risk of overfitting and simplifying the network.

4.4 Auxiliary Classifiers: GoogleNet introduced auxiliary classifiers in intermediate layers during training, encouraging the network to learn useful features at different depths and combat the vanishing gradient problem.

AlexNet was designed to take advantage of parallel processing capabilities, particularly using GPUs. This enabled faster training times and made it more feasible to train deep neural networks on large datasets. GoogleNet allows the network to capture multi-scale features efficiently and helps improve the representational power of the model without significantly increasing computational cost.

5. RESULT AND DISCUSSION

This article presents pneumonia and normal lung sound classification technique. As described in the proposed algorithm section, firstly the input lung sound recordings are transformed into three types of spectrogram images: log-mel spectrogram, conventional spectrogram and CQT spectrogram. The underlying variation in magnitude and frequency are captured in more efficient way by using three spectrograms instead of one. Use of one spectrogram only will not be able to capture the variations between pathological and normal lung sounds.

The ICBHI 2017 challenge database is used for the experimental evaluation consisting of respiratory sound recordings of healthy and pneumonia subjects.

The ICBHI dataset involves 126 subject recordings with 26 healthy and 100 samples with different respiratory diseases. The duration of these samples are not uniform (around 15-20 seconds), hence for spectrogram generation is performed using a sample of 7 second in both healthy and pathological signals. The collected 7 second segments of the lung sound recordings were used to create time-frequency representations in three different spectrograms: conventional spectrogram, log-Mel spectrogram, and IIR-CQT spectrogram. Figure 5.1 shows few samples of healthy and pneumonia spectrogram representations from the database. It is clearly observed that the two-dimensional representations of healthy and pathological spectrograms are different and can be exploited for the detection task

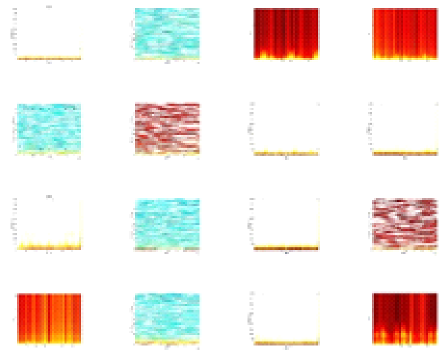

Figure 5.1 Sample spectrogram time-frequency visualization of healthy and pneumonia diseased subjects.

The proposed approach is assessed using following performance parameters:.

$$\text{F-Score} = 2 \times \frac{P \times S}{P+S}$$

Where P= precision and S =sensitivity

$$\text{F-Score} = 2 \times \frac{P \times S}{P+S}$$

Two transfer learning models, AlexNet and GoogleNet are trained using 80% of the randomly selected training images. Remaining 20% samples are used for testing the model. Figures 5.2 and 5.3 depict training plots of AlexNet and GoogleNetarchitectures respectively. The network is trained using learning rate of 0.001 and 0.0001 and 'sgdm' optimization technique. Effect of batch size on the detection rate is also analyzed in this study.

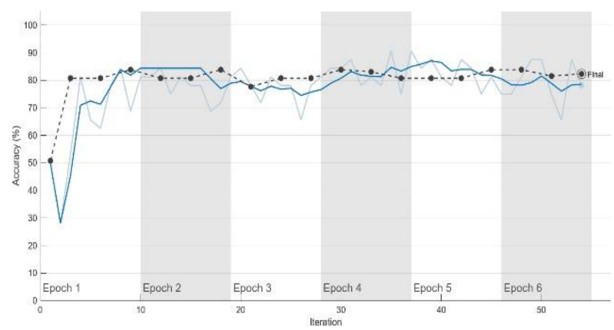

Figure 5.2 Training plot of AlexNet transfer learning model.

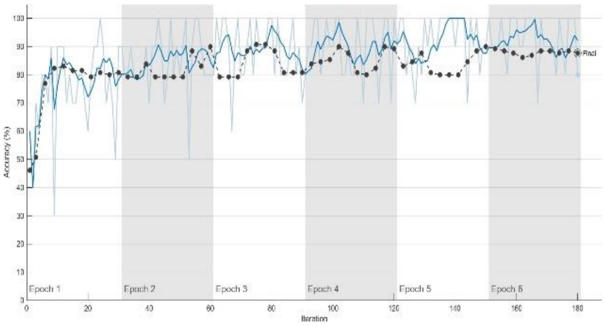

Figure 5.3 Training plot of GoogleNet transfer learning model.

Table 5.1 AlexNet F1-score and accuracy at learning rate is set at 0.0001 with different batch size.

Batch Size	F1-Score	Accuracy
16	84.86	85.24
32	81.32	81.19
64	80.00	80.47

Table 5.2AlexNet performance at learning rate =0.001.

Batch Size	F1-Score	Accuracy
16	85.16	85.38
32	81.98	82.31
64	81.74	81.54

Tables 5.1 and 5.2 shows average accuracy and F1-Score obtained using the AlexNettransfer learning model with learning rate of 0.0001 and 0.001 respectively. Highest detection rate of 85.16% is attained when the learning rate is set to 0.001 and batch size of 16. As batch size increases, the detection accuracy and F1-score decreases.

Table 5.3 Google Net F1-score and accuracy at learning rate =0.0001.

Batch Size	F1-Score	Accuracy
16	90.92	91.58
32	89.48	90.21
64	86.26	86.90

Table 5.4 GoogleNet performance at learning rate =0.001

Batch Size	F1-Score	Accuracy
16	93.86	94.64
32	91.73	92.26
64	89.62	90.48

In the second set of experiment results, GoogleNet transfer learning model is trained and tested. Tables 5.3 and 5.4 depict detection accuracy and F1-score obtained using GoogleNetusing 0.001 and 0.0001 learning rate respectively. Compared to AlexNet model, GoogleNet attained higher accuracy and F1-score. Highest detection rate of 94.64% is achieved using GoogleNet.

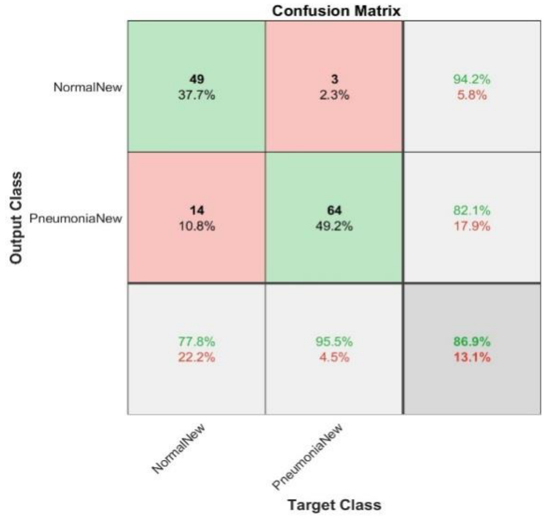

Figure 5.4 Confusion matrix using GoogleNet model at learning rate of 0.0001 and batch size of 64.

Similar pattern is also observed in GoogleNet model regarding the batch size effect on the average detection accuracy. Sample confusion matrix obtained using the GoogleNetframework is shown in figure 5.4.

Overall, the GoogleNet model produces improved detection accuracy and F1-score as compared to ALexNet. This is observed in different batch sizes and learning rate. Besides, learning rate of 0.001 is optimal for the proposed algorithm which generates best detection rate. Sample output of the deep transfer learning models depicted in figure 5.5.

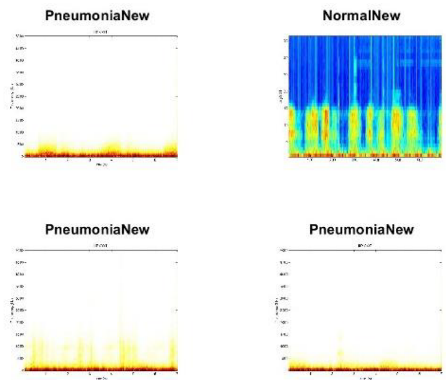

Figure 5.5. Sample outcome of the proposed deep transfer learning algorithm.

In future, the deep transfer learning based feature extraction and machine learning based classification can be explored. Additionally, the deep neural network model can be optimized to select optimal hyperparameter values automatically. Future research can also explore how the deep transfer learning approach can be tailored to accommodate a spectrum of lung pathologies, including pneumonia, chronic obstructive pulmonary disease

(COPD) and interstitial lung diseases. Future work can also address practical challenges related to deployment, usability, and integration with existing healthcare workflows.

6. CONCLUSION

This article presents an approach in the field of lung sound analysis through the application of deep transfer learning techniques. The proposed deep transfer learning model demonstrated the potential to achieve accurate and good detection accuracy of respiratory conditions. The ability to discern subtle variations in lung sounds, which are indicative of various respiratory disorders, can effectively exploited by AlexNet and GoogelNet models. As revealed from the experimental evaluations, GoogleNet attained highest detection accuracy at learning rate of 0.001 and a batch size of 16. By combining modern deep learning frameworks with clinical expertise, early detection and diagnosis of respiratory conditions can be developed enhancing the quality of life for individuals globally.

REFERENCES

[1] Kaushal B, Raveendran S, Patil MD, Birajdar GK. Spectrogram image textural descriptors for lung sound classification. Machine learning and deep learning in efficacy improvement of healthcare systems. CRC Press. 2022 May 18:109-36.

[2] Neili Z, Fezari M, Redjati A. ELM and K-nn machine learning in classification of Breath sounds signals. Int J ElectrComput Eng. 2020 Aug 1;10(4):3528-36.

[3] Aykanat M, Kılıç Ö, Kurt B, Saryal S. Classification of lung sounds using convolutional neural networks. EURASIP Journal on Image and Video Processing. 2017 Dec;2017(1):1-9.

[4] Chambres G, Hanna P, Desainte-Catherine M. Automatic detection of patient with respiratory diseases using lung sound analysis. In2018 International Conference on Content-Based Multimedia Indexing (CBMI) 2018 Sep 4 (pp. 1-6). IEEE.

[5] Jayalakshmy, S., Sudha, G.F. GTCC-based BiLSTM deep-learning framework for respiratory sound classification using empirical mode decomposition. Neural Computingand Applications, 33, 17029–17040(2021). https://doi.org/ 10.1007/s00521-021-06295-x

[6] Alqudah, A.M., Qazan, S. & Obeidat, Y.M. Deep learning models for detecting respiratory pathologies from raw lung auscultation sounds. Soft Comput 26, 13405–13429(2022).https://doi.org/10.1007/s00500-022-07499-6

[7] Dubey, R., Bodade, R.M. & Dubey, D. Efficient classification of the adventitious sounds of the lung through a combination of SVM-LSTM-Bayesian optimization algorithm with features based on wavelet bi-phase and bi-spectrum. Res. Biomed. Eng. 39, 349–363 (2023). https://doi.org/10.1007/s42600-023-00270-2

[8] Baghel, N., Nangia, V. & Dutta, M.K. ALSD-Net: Automatic lung sounds diagnosis network from pulmonary signals. Neural Comput&Applic 33, 17103–17118 (2021). https://doi.org/ 10.1007/s00521-021-06302-1

[9] L. Pham, H. Phan, R. Palaniappan, A. Mertins and I. McLoughlin, "CNN-MoE Based Framework for Classification of Respiratory Anomalies and Lung Disease Detection," in IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 8, pp. 2938-2947, Aug. 2021, doi: 10.1109/JBHI.2021.3064237.

[10] T. Nguyen and F. Pernkopf, "Lung Sound Classification Using Snapshot Ensemble of Convolutional Neural Networks," 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 2020, pp. 760-763,

[11] Ma, Yi, Xinzi Xu, and Yongfu Li. "LungRN+ NL: An improved adventitious lung sound classification using non-local block resnet neural network with mixup data augmentation." In Interspeech, pp. 2902-2906. 2020.

[12] A. Yadav, M. K. Dutta and J. Prinosil, "Machine Learning Based Automatic Classification of Respiratory Signals using Wavelet Transform," 2020 43rd International Conference on Telecommunications and Signal Processing (TSP), Milan, Italy, 2020, pp. 545-549, doi: 10.1109/TSP49548.2020.9163565.

[13] Kaushal B, Raveendran S, Patil MD, Birajdar GK. Deep Autoencoder Neural Networks for Heart Sound Classification. In Artificial Intelligence in Medical Virology 2023 Apr 22 (pp. 165-189). Singapore: Springer Nature Singapore.