Chronic Obstructive Pulmonary Disease Severity Classification using lung sound

1st Khanaghavalle G R
Computer Science and Engineering
Sri Venkateswara College of Engineering
Chennai, India
khanaghavalle@gmail.com

2nd Rahul G
Computer Science and Engineering
Sri Venkateswara College of Engineering
Chennai, India
rahulgilakareddy@gmail.com

3rd Senajith S R
Computer Science and Engineering
Sri Venkateswara College of Engineering
Chennai, India
senajithsrs@gmail.com

4th Vishnuvasan T S
Computer Science and Engineering
Sri Venkateswara College of Engineering
Chennai, India
vishnuvasants@gmail.com

5th Keerthana S
Computer Science and Engineering
Sri Venkateswara College of Engineering
Chennai, India
keerthanasrikanth.ks@gmail.com

***Abstract*—A global health concern, chronic obstructive pulmonary disease (COPD) demands early detection and intervention for effective treatment. Leveraging the extensive 12-channel lung sound dataset, RespiratoryDatabase@TR, our study establishes a robust multi-class COPD severity diagnostic system. We employ a rigorous feature extraction procedure, including spectrogram, Mel spectrogram, and chromogram analysis, alongside specific data preprocessing and augmentation methods. The RESNET50 model is chosen for training, ensuring precise classification across COPD severity levels. Our findings underscore the significance of sound-based prognoses, particularly in early COPD diagnosis. With an estimated 251 million people globally affected by COPD, innovative solutions like ours are crucial. The amalgamation of extensive datasets and advanced machine learning techniques holds the promise of transforming COPD diagnosis and treatment on a global scale, improving the lives of millions affected by this illness.**

***Index Terms*—COPD, RESNET, Deep Learning**

I. Introduction

TChronic Obstructive Pulmonary Disease (COPD) stands as a formidable global health challenge, demanding heightened attention due to its escalating prevalence and significant impact on public health. This debilitating respiratory condition, characterized by persistent airflow limitation and associated with an abnormal inflammatory response of the lungs to noxious particles or gases, has emerged as a leading cause of morbidity and mortality worldwide [1]. With an estimated 251 million individuals affected globally and an increasing prevalence, COPD has transcended regional boundaries to become a critical public health concern [9].

Early detection and intervention are paramount for effective COPD management, as timely diagnosis can not only enhance the quality of life for affected individuals but also mitigate the burden on healthcare systems. However, diagnosing COPD in its incipient stages remains a challenging task, often leading to delayed intervention and compromised treatment outcomes. In light of this, there is a growing recognition of the need for innovative diagnostic tools that can facilitate early identification and classification of COPD severity, paving the way for personalized and targeted therapeutic interventions. The advent of digital health technologies, coupled with the increasing availability of large-scale medical datasets, has ushered in a new era of possibilities for respiratory disease diagnostics. Among the myriad of approaches, the utilization of lung sound analysis presents a promising avenue for non-invasive and accessible COPD severity assessment. Lung sounds, generated by the turbulent flow of air through the respiratory system, carry valuable information about the underlying pulmonary condition [5]. Analyzing these acoustic signals can provide insights into respiratory abnormalities, offering a potential solution for early COPD detection. In response to this imperative, our study harnesses the power of the extensive Respiratory-Database@TR, a comprehensive 12-channel lung sound dataset, to develop a robust multi-class COPD severity diagnostic system. The richness of this dataset allows for a nuanced exploration of the diverse manifestations of COPD, enabling a more accurate and granular classification of disease severity. Through a meticulously designed methodology, we embark on a journey of feature extraction, leveraging spectrogram, Mel spectrogram, and chromogram analyses. These techniques capture the intricate patterns within lung sounds, laying the foundation for a sophisticated diagnostic framework.

The significance of data preprocessing and augmentation in enhancing the performance of machine learning models cannot be overstated. In our pursuit of a reliable diagnostic system, we employ extensive preprocessing methods to ensure the integrity and quality of the dataset. Augmentation techniques are then applied judiciously, enriching the dataset with diverse instances and fortifying the model against potential biases. The culmination of these steps sets the stage for the application of the RESNET50 model, a robust deep-learning architecture renowned for its efficacy in image recognition tasks.

Our choice of the RESNET50 model is deliberate, guided by its ability to discern intricate patterns and relationships within complex datasets. Transcending traditional image classification, the RESNET50 model demonstrates its versatility in handling the multifaceted nature of lung sound data. Through a rigorous training process, our model becomes adept at classifying COPD severity levels with a level of precision that holds promise for clinical applications.

This study underscores the pivotal role of sound-based prognoses in the identification of respiratory diseases, with a specific focus on the challenges and opportunities presented by early COPD diagnosis. As we delve into the intricacies of our methodology and the outcomes of our analysis, it becomes evident that the fusion of extensive datasets and cutting-edge machine learning techniques has the potential to redefine the landscape of COPD diagnostics.

The urgency of our research is underscored by the staggering prevalence of COPD globally. With an estimated 251 million people grappling with the repercussions of this disease, innovative solutions are not just desirable; they are imperative. Our work represents a step towards addressing this substantial public health issue, offering a glimmer of hope for individuals affected by COPD and heralding a paradigm shift in the way this condition is diagnosed and treated on a global scale. As we navigate the complexities of respiratory health, the marriage of advanced technologies and comprehensive datasets emerges as a beacon of progress, promising to enhance the lives of millions affected by the insidious grip of COPD

II. Related works

Arka et al. [10] utilized YAMNet architecture to classify the COPD levels. The dataset was augmented and the Mel spectrogram feature was extracted from the lung sounds. Oweis et al. [11] performed binary classification on respiratory sound using artificial neural networks. The work utilized an autocorrelation function in the feature extraction stage to perform the classification. Jayalakshmy et al. [7] used the ALEXnet architecture to perform four classifications on lung sound. Scalogram feature was extracted from the sound and various optimization techniques which shall be used are examined. Pham et al. [7] explored a robust deep neural network for lung sound analysis. They have extracted the spectrogram feature from the lung sound and proposed a teacher-student scheme to reduce the model complexity. Fernandez-Granero et al. [8] created their own dataset to analyse the main cause of the disease COPD in patients. The feature extracted was a discrete wavelet transform along with a correlation filter. The decision-based classifier was used to do the binary classification of the disease. Altan et al. [3] utilised an ensemble machine learning classifier for classifying the COPD severities. Altan et al. [4] explored second-order difference plot features for the non-linear analysis of the lung sound. The work used s deep belief network to perform the severity classification on COPD levels. Naves et al. [6] utilized a genetic algorithm to classify lung sounds. Higher-order feature extraction techniques were deployed to the lung sounds.

III. Research Methodology

The digital stethoscope adeptly captures pivotal audio data formatted in the '.wav' file, spanning a substantial 17-second interval. This critical information is intricately integrated into the RepositoryDatabase@TR. The subsequent phase entails a meticulously designed and sophisticated procedure, encompassing rigorous preprocessing and augmentation methodologies. These meticulous steps are precisely tailored to refine and optimize the dataset, ensuring its readiness for subsequent utilization within our model crafted for advanced multi-class classification - low severity (COPD 0) to high severity (COPD 4).

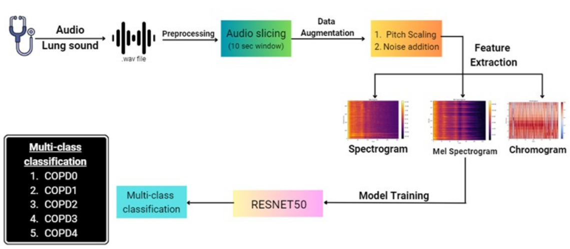

Fig. 1. Architecture of the proposed work.

A. Dataset

The dataset - RespositoryDatabase@TR [2] contains 12 distinct channels of lung sounds meticulously focused on the upper, middle, and lower lung regions, as well as the costophrenic angle areas on both the posterior and anterior sides of the chest. These recordings undergo a rigorous validation process and are meticulously labeled through the evaluation of chest X-rays, Pulmonary Function Tests (PFT), and auscultation sounds by two pulmonologists. The database comprises data from 30 healthy subjects and 45 subjects diagnosed with pulmonary diseases, including asthma, chronic obstructive pulmonary disease, and bronchitis. The uniqueness of the database lies in its integrated assessment approach, combining auscultation sound results, chest X-rays, and PFT, thereby providing synchronous evaluation capabilities for lung sounds.

B. Preprocessing

The established average time window for each audio recording stands at 17 seconds. To ensure uniformity and facilitate normalization, a systematic segmentation approach is implemented, dividing the audio data into 10-second intervals. For example, in a 17-second recording, segments are derived from 1 to 10 seconds and 8 to 17 seconds. In instances of a 23-second recording, segments are strategically extracted from 1 to 10 seconds, 11 to 20 seconds, and 14 to 23 seconds. This also ensures overlapping of data as in the case of 17 seconds (8 to 10 seconds) and 23 seconds (14 to 20 seconds). The selected 10-second window duration is judiciously chosen, encompassing approximately two lung cycles, rendering it pivotal for thorough information analysis. Consequently, each audio data entry is precisely standardized to a 10-second duration, ensuring consistency in subsequent analysis.

C. Augmentation

Data augmentation [14] serves as a pivotal technique in machine learning, particularly in audio data analysis, mitigating the challenge of data scarcity during model training. The inherent limitation in the size of audio datasets necessitates strategic augmentation to expand the dataset through variations of existing samples. This augmentation practice significantly contributes to the model's generalization and overall performance enhancement. In this study, we leverage the robust capabilities of the Librosa package, a widely adopted Python library for audio analysis and manipulation. Renowned for its effectiveness, Librosa facilitates diverse augmentation tasks, with our focus on two critical operations: Pitch Scaling and Noise Addition. These operations are meticulously applied to enrich the dataset systematically. Librosa's comprehensive functionalities for loading, processing, and transforming audio data ensure the seamless and efficient implementation of these augmentation techniques, affirming its suitability for advanced audio dataset manipulation within the context of machine learning.

D. Feature extraction

The dataset containing audio data in '.wav' format serves as the foundational input for our research. To enable subsequent sophisticated processing, a meticulous transformation of these audio files into visual representations is undertaken. This transformation encompasses the generation of Spectrograms, Melspectrograms, and Chromograms [12] from the raw audio data. These image representations encapsulate indispensable features vital for comprehensive downstream analysis. Subsequently, the derived images are seamlessly integrated into our model architecture, purposefully designed to comprehend and analyze audio data in a refined 2D image format. This strategic transition from audio to image data significantly augments the model's capacity to discern intricate patterns and features embedded within the audio signals. Notably, Spectrograms offer a comprehensive visual representation of frequency content over time, while Melspectrograms emphasize perceptually relevant frequencies, and Chromograms focus on pitch and harmonic aspects, collectively enhancing the interpretability of audio data for advanced signal processing and machine learning applications.

E. RESNET model

The incorporation of the ResNet-50 [13] architecture significantly enhances the depth and complexity of our model for audio data analysis. ResNet-50 is a convolutional neural network (CNN) known for its deep residual learning characteristics, incorporating skip connections to mitigate the vanishing gradient problem. In our model, the ResNet-50 architecture functions as a feature extractor, capturing intricate hierarchical features present in the transformed audio images. The model's parameters, including convolutional kernel sizes, filter depths, and fully connected layers, are fine-tuned through iterative optimization processes. The deep structure of ResNet-50 facilitates the learning of complex patterns inherent in audio representations, contributing to improved classification performance. The integration of ResNet-50 into our architecture is a judicious choice to enhance the model's capacity for discerning subtle features in the context of audio signal analysis.

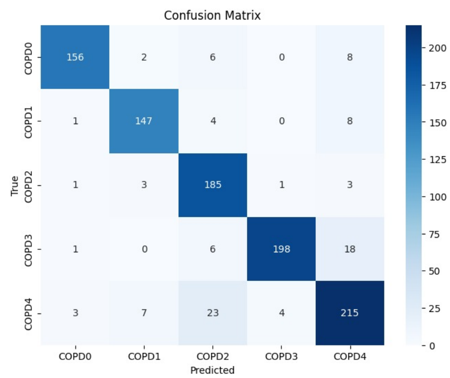

Fig. 2. Confusion matrix for the classification

F. Experimental results

The multi-class classification performance of the COPD severity levels using precision, recall, and F1-score metrics is highly commendable. Class COPD0 demonstrates exceptional precision (0.96), indicating a minimal false-positive rate in identifying individuals with mild COPD. COPD3 exhibits notable precision (0.98), signifying accurate identification of severe cases. COPD2 displays balanced performance with high recall (0.96), implying effective recognition of moderate cases. COPD4, although displaying slightly lower precision, maintains respectable scores across precision (0.85), recall (0.85), and F1-score (0.85), indicating consistent performance across the spectrum. As a result, the model demonstrates robust capability in discerning COPD severity, showcasing particularly high precision in milder cases and reliable identification of severe instances. These results affirm the model's potential for precise clinical applications in classifying diverse COPD severity levels, underlining its significance in advancing respiratory disease diagnostics and management.

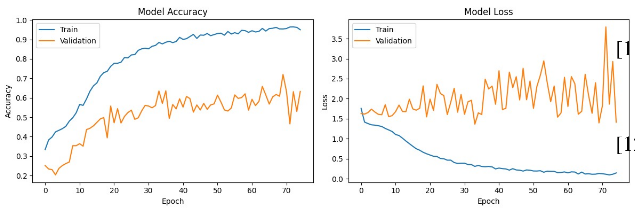

Fig. 3. Accuracy and loss plot for the classification

IV. Conclusion and Future Works

Our study introduces a 2 dimensional image centric audio embedded snippet-based framework for COPD severity classification, utilizing time-frequency representations of lung sounds and a ResNet50-based transfer learning model. The framework exhibits promising results, achieving an accuracy of 94.57

References

[1] Adeloye, Davies, et al. ”Global and regional estimates of COPD prevalence: Systematic review and meta–analysis.” Journal of global health 5.2 (2015).

[2] G. Altan, Y. Kutlu, Y. Garbi, A. Ö. Pekmezci, and S. Nural, “Multimedia respiratory database (Respiratory-Database@TR): Auscultation sounds and chest X-rays,” Natural Eng. Sci., vol. 2, no. 3, pp. 59–72, Oct. 2017.

[3] G. Altan, Y. Kutlu, and A. Gökçen, “Chronic obstructive pulmonary disease severity analysis using deep learning on multi-channel lung sounds,” TURKISH J. Electr. Eng. Comput. Sci., vol. 28, no. 5, pp. 2979–2996, Sep. 2020

[4] Altan, Gokhan, et al. ”Deep learning with 3D-second order difference plot on respiratory sounds.” Biomedical Signal Processing and Control 45 (2018): 58-69.

[5] Naqvi, Syed Zohaib Hassan, and Mohammad Ahmad Choudhry. ”An automated system for classification of chronic obstructive pulmonary disease and pneumonia patients using lung sound analysis.” Sensors 20.22 (2020): 6512.

[6] R. Naves, B. H. G. Barbosa, and D. D. Ferreira, “Classification of lung sounds using higher-order statistics: A divide-and-conquer approach,” Comput. Methods Programs Biomed., vol. 129, pp. 12–20, Jun. 2016.

[7] R. Oweis, E. Abdulhay, A. Khayal, and A. Awad, “An alternative respiratory sounds classification system utilizing artificial neural networks,” Biomed. J., vol. 38, no. 153, p. e61, 2015.

[8] L. Pham, H. Phan, R. Palaniappan, A. Mertins, and I. McLoughlin, “CNN-MoE based framework for classification of respiratory anomalies and lung disease detection,” IEEE J. Biomed. Health Informat., vol. 25, no. 8, pp. 2938–2947, Aug. 2021.

[9] Quaderi, S. A., and J. R. Hurst. ”The unmet global burden of COPD.” Global health, epidemiology and genomics 3 (2018): e4.

[10] Roy, Arka, and Udit Satija. ”A Novel Melspectrogram Snippet Representation Learning Framework for Severity Detection of Chronic Obstructive Pulmonary Diseases.” IEEE Transactions on Instrumentation and Measurement 72 (2023): 1-11.

[11] S. Jayalakshmy and G. F. Sudha, “Scalogram based prediction model for respiratory disorders using optimized convolutional neural networks,” Artif. Intell. Med., vol. 103, Mar. 2020, Art. no. 101809.

[12] Wei, Shengyun, Shun Zou, and Feifan Liao. ”A comparison on data augmentation methods based on deep learning for audio classification.” Journal of physics: Conference series. Vol. 1453. No. 1. IOP Publishing, 2020.

[13] Wightman, Ross, Hugo Touvron, and Hervé Jégou. ”Resnet strikes back: An improved training procedure in timm.” arXiv preprint arXiv:2110.00476 (2021).

[14] Wei, Shengyun, Shun Zou, and Feifan Liao. ”A comparison on data augmentation methods based on deep learning for audio classification.” Journal of physics: Conference series. Vol. 1453. No. 1. IOP Publishing, 2020.