Advanced Deep Learning Techniques for Lung Sound Classification: Binary, Multi-Class, and Ensemble Approach

Gayan Perera
Department of Computer and Data Science
NSBM Green University
Homagama, Sri Lanka
gayanp@nsbm.ac.lk

Heshan Chandeepa Pathmakumara
Department of Computer and Data Science
NSBM Green University
Homagama, Sri Lanka
bhcpathmakumara@students.nsbm.ac.lk

2025 7th International Conference on Software Engineering and Computer Science (CSECS) | 979-8-3315-2221-6/25/$31.00 ©2025 IEEE | DOI: 10.1109/CSECS64665.2025.11009363

***Abstract*— Detecting respiratory diseases, particularly lung disorders is crucial due to their global prevalence. Auscultation by experts has been the standard method for detecting lung diseases. However, the efficiency of diagnosis can be greatly increased by automating this procedure. Using Convolutional Neural Networks (CNN) for binary classification and a CNN, CNN-LSTM hybrid model for multi class classification, this study uses deep learning approaches to automate lung sound classification. To increase classification accuracy, the study extracts important elements from lung sounds such as chromograms, Mel spectrograms and Mel Frequency Cepstral Coefficients (MFCCs). Ensemble learning which combines CNN and CNN-LSTM for multi-class classification, substantially improves the model. With an accuracy of 86.11% the suggested ensemble model surpasses current models in the field. The accuracy of binary classification is 88.38%. This study demonstrates the advantages of combining multiple architectures to improve classification performance and shows how deep learning models can be used to effectively classify lung sounds. The results have significant implications for the early detection of lung diseases.**

Keywords—artificial intelligence, ensemble learning, respiratory diseases, sound processing

I. Introduction

Respiratory disorders are a significant worldwide health concern, contributing to high morbidity and mortality rates. Respiratory diseases are an important health issue worldwide. As of 2017, lung disorders affected around 545 million people globally, and their frequency has increased by 40% since 1990, making them a serious global health concern. Chronic respiratory conditions such as asthma and chronic obstructive pulmonary disease (COPD) are the third leading cause of death globally, after diseases such as cardiovascular disease and cancer. Despite medical advances, many people continue to receive poor treatment and delayed diagnoses, especially in areas with underfunded medical systems. Risk factors like smoking, air pollution are responsible for high rates of disability and mortality [1].

In Sri Lanka also, the burden of chronic respiratory diseases, specifically COPD is significant with a spread rate of nearly 10.5%. The lack of structured pulmonary rehabilitation (PR) services and lack of knowledge among healthcare professionals regarding the benefits of PR further complicate to manage these conditions [2].

Traditional diagnostic methods, mostly manual auscultation, rely primarily on the knowledge of health care professionals. Not only is this procedure very time-consuming, but it also varies depending on the experience of the physician, the quality of the stethoscope, and external environmental factors such as background noise [3] ,[4]. There are three main categories of lung sounds. They are Wheeze, Crackles and Rhonchi. Those sounds identification of the difference between those sounds can be challenging to clinicians, especially in environments with overlapping sounds such as heartbeats or coughing [5]. The above barriers can lead to delayed diagnosis, which has an impact on patient outcomes. This can lead to inconsistent or incorrect diagnoses, especially in noisy settings or when doctors are inexperienced.

In recent years, the diagnosis of diseases such as neurological diseases, cancers and respiratory disorders from AI has made a huge impact in the healthcare industry [6], [7], [8]. Lung sound classification has emerged as a promising method for automating respiratory diagnosis with the advent of deep learning (DL) techniques in healthcare, particularly Convolutional Neural Networks (CNNs). CNNs are more efficient at extracting features from lung sound spectrograms. Even though they have been shown to be successful, the majority of recent studies have mostly concentrated on binary or small multi-class classifiers, training models on a small number of patient classes [9], [10]. The limited number of disease classes in the datasets used in the studies cited here means that they do not adequately address the diversity of diseases. This is a fundamental problem facing clinicians, and these studies provide insightful information about the applications of AI in healthcare. One of the least studied topics in the field of lung sound classification is the use of Ensemble Learning techniques in the prediction of classification models. Our work primarily addresses the problem of lung sound classification in ten disease classifications as a new addition to previous research. Additionally, we emphasized how

Authorized licensed use limited to: University of Plymouth. Downloaded on November 24,2025 at 09:58:19 UTC from IEEE Xplore. Restrictions apply.

combining stacked audio-specific features instead of separate features to input the CNN models we use here increases model performance.

The following are the primary contributions of the paper.

- To increase the range of diseases addressed, we propose to combine two publicly available datasets, namely the ICBHI 2017 Challenge dataset [11] and the dataset created by Fraiwan et al.[12].
- Our paper proposes to combine three types of features, Mel Spectrogram, MFCC, and Chromagram, rather than using only one for each audio sample. This method can capture a wide range of features from lung sounds by using a 3D representation of the features to improve the performance of the model.
- Develop a binary classification using CNN to classify whether the audio file is a healthy one or abnormal one.
- Develop a multi class classification using CNN and CNN-LSTM hybrid model to classify 10 classes of lung diseases.
- Develop an Ensemble Learning model using CNN model and the CNN-LSTM hybrid model.

Our paper is organized as follows. Section II describes the literature review on lung sound classification for respiratory disease detection. In Section III, we delve into the methodology of the developed models and ensemble learning technique. Section IV analyses the results of the classification models. And in the Section V, discuss about the existing multi class classification model from the existing research. Next Section VI is the conclusion of the study.

II. Background

In DL-based models, the dataset plays a crucial role in building an optimal solution. In the domain of lung sound classification, two major publicly available datasets stand out. The ICBHI 2017 [11] Challenge dataset includes 920 sound recordings across 8 classes, including the “healthy” class. The dataset by Fraiwan et al.[12] originally contained 112 recordings, and was expanded to 336 after augmentation, covering 11 respiratory disease categories, including the “healthy” class.

Kim et al. [13] collected 1,918 respiratory sound recordings (normal, crackles, wheezes, and rhonchi) from a clinical context and applied pretrained image feature extraction with a CNN classifier. They have used basic categories of abnormal lung sounds. In study [14], Used a dataset of children's lung sounds, the researchers created a DenseNet169 CNN model using efficient preprocessing methods. In this study [15], they classify respiratory sound anomalies such as wheezing and crackles using ICBHI 2017 dataset. Shuvo et al. [16] for five respiratory disease classes.

Researchers used various augmentation strategies applied to ensure uniformity between dataset classes. Acharya et al. [15] used pitch shifting, random shifting, speed variation, and noise addition to provide improved samples in order to solve the lack of lung sound recording data and data imbalance. Truc et al. [17] to increase the model's attention to minority classes,vocal tract length perturbation (VTLP), temporal stretching, and the focal loss objective were applied.

Extracting audio features and feeding them into classifiers improves classification accuracy. In this study [18], The researchers preprocess lung sound signals to remove noise, transform them into spectrogram using short-time Fourier transforms, and classify these spectrogram with a deep learning network based ResNet. Zeenat et al.[9] have created a CNN model with a classification accuracy of 99% using image feature vectors collected from spectroscopic, mel-frequency cepstral coefficients (MFCC) and chromogram features. It is a model classified into six noise classes. In here [17], the authors of the research have developed a robust lung sound classification system using snapshot ensembles of Convolutional Neural Networks (CNNs). It also extracts high-level features from the log mel-spectrogram using a CNN architecture.

Furthermore, ensemble learning has been successfully used to improve diagnostic robustness and accuracy in lung sound classification. This Study [17], used an ensemble of convolutional neural networks (CNNs) snapshots to propose a reliable and effective lung sound classification system, demonstrating better performance than individual models.

Kim et al. [19] also presented an ensemble learning model for classifying respiratory abnormalities such as normal, crackles, wheezes, and rhonchi. The model demonstrated the potential of ensemble approaches in respiratory sound analysis by successfully distinguishing between several lung sound categories through the integration of multiple classifiers.

Furthermore, previous methods often rely on single feature representations when extracting features from audio data. To improve the model’s capacity to recognize complex patterns, we use a stacked combination of three audio features in this study. CNN is used for binary classification, and a CNN and CNN-LSTM hybrid model is used for multi-class classification. The strengths of the CNN and CNN-LSTM models are finally combined in an ensemble learning strategy, demonstrating increased accuracy and resilience in lung sound classification tasks.

III. Methodology

A. Process Overview

In this study, Deep learning (DL) approaches have been used to ensure accurate and robust disease detection for this proposing pipeline. It integrates standardization through resampling, segmentation, normalization, preprocessing, feature extraction, model development and classification to ensure the consistency of the dataset. As shown in Figure 1, the process flow consists of three deep learning models. A CNN for binary classification for healthy vs abnormal sounds, a CNN multiclass model and a hybrid model combining convolutional and recurrent layers for multiclass classification of specific lung diseases.

Authorized licensed use limited to: University of Plymouth. Downloaded on November 24,2025 at 09:58:19 UTC from IEEE Xplore. Restrictions apply.

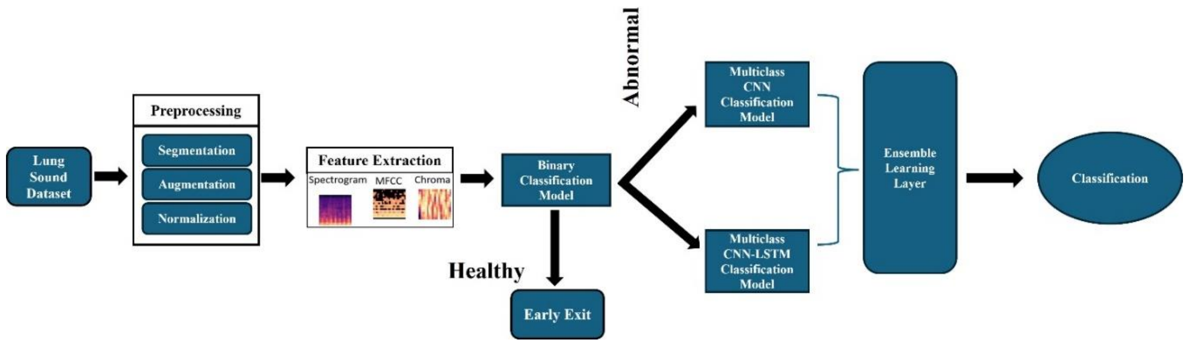

Fig. 1. Overall process

B. Data Preprocessing and Feature Extraction

In this study we combined two publicly accessible resources. The datasets are the ICBHI 2017 dataset [11] and the dataset developed by Fraiwan et al [12]. There were data imbalance in the merged dataset, and heart failure was linked to certain disease categories. In order to train classification models, we choose nine samples with abnormal conditions and healthy samples. As the primary step, the audio segmentation process is done. For further analysis and to standardize the dataset, the lung sound recordings are divided into segments of fixed length of six seconds. Using the pydub library, we processed recordings from selected classes including Asthma, Bronchiectasis, Bronchiolitis, Bronchitis, COPD, Healthy, Lung Fibrosis, Pleural Effusion, Pneumonia and URTI. Each audio file was first converted to stereo (if not already) and resampled to a fixed sampling rate of 44,100 Hz to ensure uniformity. The sample width was standardized to 16-bit for numerical stability during processing. 16-bit audio is computationally efficient and has a wide enough dynamic range to record respiratory sounds. Then the data augmentation procedure was implemented to resolve class imbalance and improve the model's ability to accurately detect lung disease. Pitch shifting, a technique that changes the pitch of audio recordings, was used to present differences to underrepresented classes while maintaining their essential qualities. This method shifts the pitch by one semitone and is called via the librosa.effects.pitch_shift library. Increasing the number of audio samples in less-growing disease classes to 200 per class produced a balanced dataset for reliable training. Furthermore, the total number of samples in the healthy class was increased to 1,900, which is approximately equal to the sum of samples in all unhealthy classes (1,800). This reduced the potential for bias during model training by ensuring equivalence between classes.The model's ability to generalize across different lung sound patterns and its classification accuracy for healthy and diseased conditions are enhanced by consistently representing all conditions in the balanced dataset. Audio data are scaled to a range between -1 and 1 using the min max scaling technique that allows for normalization. By minimizing amplitude variations across the dataset, this modification improves the numerical stability and convergence of training.The librosa library was used to load the raw audio data from each audio file.

The next step involved the feature extraction process. Three types of features were extracted. They are MFCC (Mel Frequency Cepstral Coefficients), Mel-Spectrogram, and Chromagram, which were stacked to form a 3D representation for each audio sample. With its 128 coefficients, MFCCs provide a concise representation of the frequency content of audio by encoding the timbral and perceptual characteristics of lung sounds. Standardized parameters, such as a hop length of 512, a Fast Fourier Transform (FFT) size of 2048, and a set duration of 264 frames for temporal consistency, were used to extract features in order to guarantee compatibility. The output 3D representation ($128 \times 264 \times 3$) offers a rich input structure by stacking features along the third axis, guaranteeing that the model obtains thorough and complementary information for reliable training.

C. Model Development

Developing a binary classification model to differentiate between normal and abnormal lung sounds is the first stage in our lung sound classification process. By facilitating a two-stage classification pipeline, this method allows the system to recognize healthy situations and terminate early without requiring additional processing.

We proposed a customized CNN as the binary classifier of this study. The model architecture used to extract the hierarchical feature representation from the input consists of 5 convolutional layers and max-pooling layers. To reduce overfitting, a solid layer with 512 units and a dropout rate of 0.5 is added. We used Rectified Linear Units (ReLU) as the activation function between layers. A sigmoid activation function is used in the output layer, and Adam optimization with a learning rate of 10^{-4} is used to optimize the model.Correct predictions for the two classes are ensured by the binary cross-entropy loss function. The model was configured to train for 100 epochs. The architecture is shown in Figure 2.

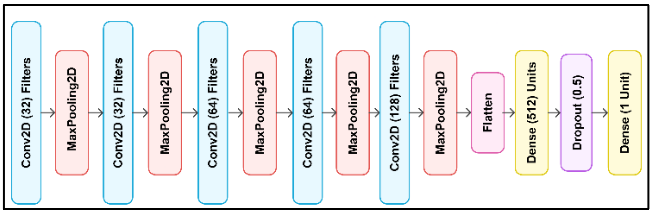

Fig. 2. Binary Classification Model

If the binary model detects aberrant lung sounds, we build a multi-class classification CNN model for further investigation. URTI, bronchiectasis, bronchiolitis, bronchitis, COPD, lung fibrosis, pleural effusion, pneumonia, and asthma are the nine lung disease categories that the model predicts. Five transform and maximum pooling layers form the architecture, and the output layer for the nine classes is softmax enabled. With a learning rate of 10^{-4} and the Adam optimizer, the sparse classifier handles the cross-entropy

Authorized licensed use limited to: University of Plymouth. Downloaded on November 24,2025 at 09:58:19 UTC from IEEE Xplore. Restrictions apply.

multi class nature well. The detailed model architecture is shown in Figure 3.

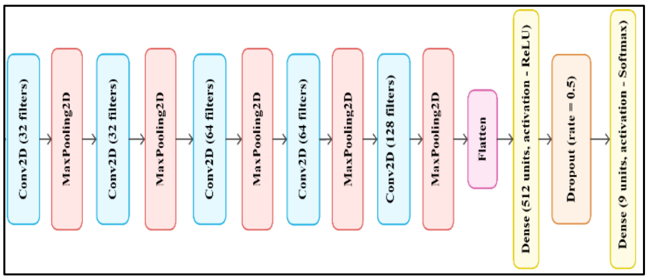

Fig. 3. Multi class classification CNN model architecture

We then build a hybrid CNN-LSTM model, which blends the sequential modeling power of LSTM with the spatial feature extraction capabilities of CNN. While the LSTM layers record the temporal relationships in lung sounds, the CNN component takes local information out of the spectrograms. Because lung sounds frequently show dynamic patterns over time, this hybrid architecture works especially well with time series data. The hybrid model can analyze both spatial and temporal data since it consists of three convolutional layers followed by LSTM layers. The output layer employs the same nine-class configuration and softmax activation as the standalone CNN model. Figure 4 is the model architecture.

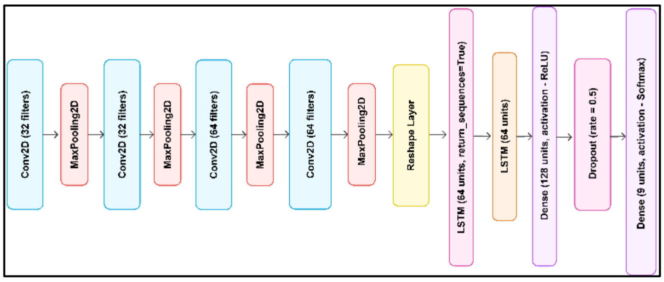

Fig. 4. Multi class classification Hybrid model (CNN-LSTM)

CNN is useful for locating local frequency patterns since it concentrates on obtaining spatial characteristics from lung sound spectrograms. By adding temporal dependencies, CNN-LSTM, on the other hand, expands on this feature and enables the model to record consecutive changes in respiratory sounds, which are essential for precise classification.

To improve classification accuracy and robustness, an ensemble learning approach was used in the final step of model building. The CNN and CNN-LSTM hybrid model outputs were blended using a weighted averaging approach, which assigned weights of 0.4 and 0.6, respectively. The hybrid model's greater accuracy and capacity to capture both spatial and temporal data is reflected in its higher weights. The ensemble system improves performance and generalization by combining predictions from both models, ensuring more accurate lung sound classification for real-world medical applications.

IV. Results Analysis

The results of our proposed models demonstrate their efficiency and performance in lung sound classification. The binary CNN model effectively discriminates between normal and abnormal lung sounds using stacking of audio features such as Mel spectrogram, MFCC, and Chromagram. Several lung disorders can be efficiently classified using a multi-class CNN model using the same feature stacking. A hybrid CNN-LSTM model was designed to enhance the performance and multi-class predictions were further improved by ensemble learning techniques. For each model, confusion matrix analysis and the obtained precision, accuracy, recall, and F1-score values are shown in Table I. To examine the performance by class, we used a confusion matrix that graphically represents the efficiency of the classification method. The confusion matrix for the binary classification model, shown in Figure 5, shows that both classes performed well overall in terms of accuracy.

TABLE I. RESULTS OF CLASSIFICATION.

Model	Accuracy	Precision	Recall	F1Score
Binary	88.38%	0.88	0.88	0.88
Multi- CNN	84.72%	0.86	0.85	0.85
Muti-CNN LSTM	83.06%	0.84	0.83	0.83
Ensemble	86.11%	0.87	0.86	0.86

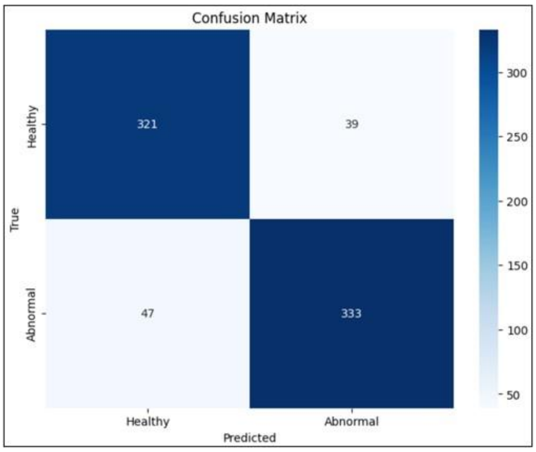

Fig. 5. Confusion matrix for binary classifier

The proposed multi-class CNN model performed well in detecting a range of lung conditions. Asthma, bronchiectasis, bronchiolitis, bronchitis, COPD, lung fibrosis, pneumonia, and URTI are represented by classes 0 to 9 in the confusion matrix for the analysis of the results shown in Figure 6.

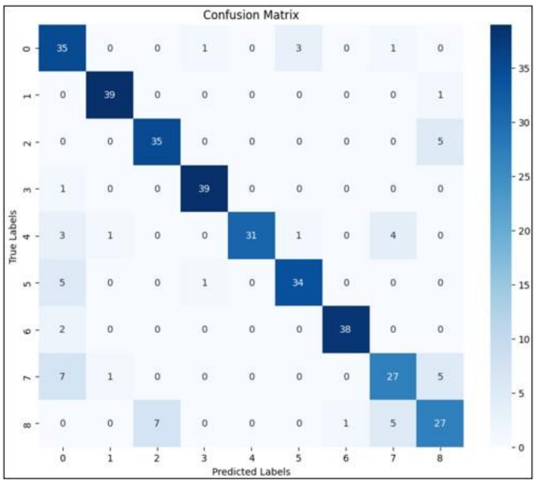

Fig. 6. Confusion matrix for multi-class CNN classifier

Authorized licensed use limited to: University of Plymouth. Downloaded on November 24,2025 at 09:58:19 UTC from IEEE Xplore. Restrictions apply.

The findings of the proposed hybrid multi-class model that combines CNN and LSTM layers demonstrate the effectiveness in classifying lung disorders. The confusion matrix in Figure 7 provides a comprehensive visual representation of the model's predictions, along with information about how well it performed for each class. The categories from 0 to 9 include pulmonary fibrosis, pneumonia, pleural effusion, URTI, COPD, bronchiectasis, bronchiolitis, bronchitis, and asthma.

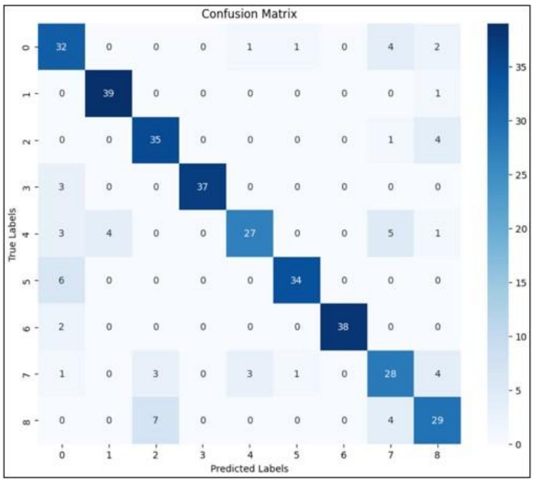

Fig. 7. Confusion matrix for multi-class CNN-LSTM

Figures 8, Figure 9 and Figure 10 show the training and loss curves on the binary classifier, multi-class CNN classifier and multi-class CNN-LSTM classifier respectively. The features and variability found in the lung sound dataset are primarily responsible for the minor spikes in the accuracy and loss curves. As can be seen from the overall decreasing trend in loss and increasing trend in accuracy, the CNN and CNN-LSTM models are successfully learning and generalizing from the stacked features.

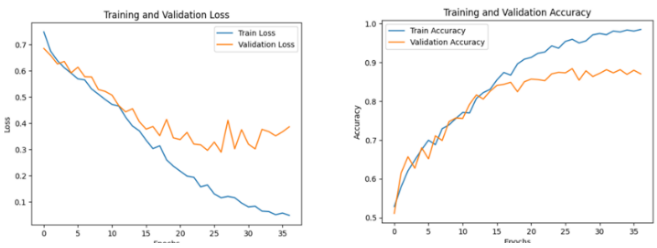

Fig. 8. Training and Loss curves for binary classifier

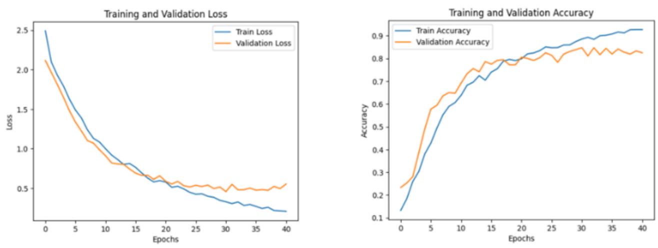

Fig. 9. Training and Loss curves for multi-class CNN classifier

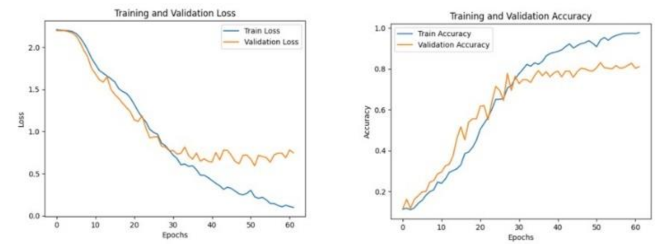

Fig. 10. Training and Loss curves for multi-class CNN-LSTM classifier

The results of the ensemble learning model show that it can correctly classify lung disorders into nine different classifications. A complete summary of the model classification performance is given by the confusion matrix, which is shown in Figure 11 and demonstrates its ability to reduce misclassifications. This shows how well the ensemble manages classes with overlapping features. This model produces robust and reliable predictions by combining the temporal learning skills of Hybrid CNN-LSTM with the spatial feature extraction capabilities of CNN highlighting the benefits of mixing complementary feature learning techniques.

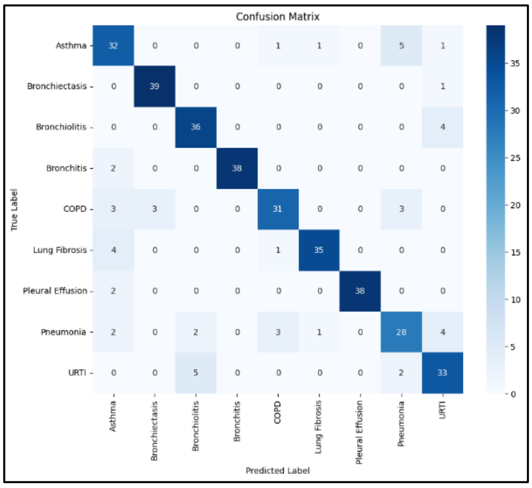

Fig. 11. Confusion matrix for multi-class ensemble classifier

To address the computational complexity, we measured the processing time for the CNN and CNN-LSTM models. In contrast to the CNN-LSTM model (62 epochs, ~66-89s/epoch), the CNN model showed faster convergence (41 epochs, ~70-87s/epoch), and showed greater efficiency while maintaining high accuracy, making it more suitable for real time applications.

V. Discussion

In this work, we introduced an ensemble learning approach and a binary classifier for lung sound classification. The ensemble model was created by combining CNN and hybrid CNN-LSTM architectures. Audio samples were transformed from the time domain into a three dimensional representation of time frequency domain features to provide input to the CNN. To provide a comprehensive representation, these features including Mel spectra, MFCC, and chromogram were stacked on top of each other. This enabled the models to efficiently learn and classify lung sound patterns.

Furthermore, we compared the results of our proposed models with those of previous multiclass classification research. Table II provides a comparison of the results with the existing studies. This comparison demonstrates the effectiveness of the ensemble learning method in resolving the difficulties related to lung sound classification highlighting the advantages and improvements of our approach in terms of accuracy, feature representation and resilience.

Authorized licensed use limited to: University of Plymouth. Downloaded on November 24,2025 at 09:58:19 UTC from IEEE Xplore. Restrictions apply.

TABLE II. COMPARISON OF EXISTING STUDIES.

Study	Model	Accuracy
[20]	CNN	76%
[21]	DS-CNN	83.46%
[22]	VGG16 CNN	66%
[23]	CNN	43%
[24]	CNN	80.4%
Our study	ENSEMBLE	86.11%

To improve its practical utility, this study can be expanded by adding more features. For lung sound classification, a desktop application can be created that provides an intuitive user interface for real time analysis and prediction. In addition, the inclusion of explainable AI (XAI) methods increases model transparency allowing medical professionals to more clearly understand the reasoning behind the predictions. These developments not only make the system more reliable and accessible, but also make it easier for clinical decision makers to use.

VI. CONCLUSION

This study successfully combines CNN and CNN-LSTM models to demonstrate how ensemble learning can advance lung sound classification. The proposed ensemble model outperformed a number of current multi-class classification methods, with a maximum accuracy of 86.11%. Furthermore, the binary classification model demonstrated its resilience in distinguishing between normal and abnormal lung sounds with an accuracy of 88.38%. The study emphasizes the importance of blending different audio features to improve classification performance by stacking MFCC, Mel Spectrogram and Chromagram as feature representations.

REFERENCES

[1] GBD Chronic Respiratory Disease Collaborators, "Prevalence and attributable health burden of chronic respiratory diseases, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017," *Lancet Respir Med*, vol. 8, no. 6, pp. 585–596, Jun. 2020, doi: 10.1016/S2213-2600(20)30105-3.

[2] U. Wijayasiri, "Developing Appropriate Pulmonary Rehabilitation Services in Sri Lanka: Assessment of People Living with COPD and Healthcare Providers in Urban and Semi Urban Areas in Sri Lanka," *International Journal of Chronic Obstructive Pulmonary Disease*, Jan. 2022, Accessed: Oct. 22, 2024. [Online]. Available: https://www.academia.edu/115588642/Developing_Appropriate_Pulmonary_Rehabilitation_Services_in_Sri_Lanka_Assessment_of_People_Living_with_COPD_and_Healthcare_Providers_in_Urban_and_Semi_Urban_Areas_in_Sri_Lanka

[3] "(PDF) Fundamentals of Lung Auscultation," *ResearchGate*, Oct. 2024, doi: 10.1056/NEJMra1302901.

[4] M. Sarkar, I. Madabhavi, N. Niranjan, and M. Dogra, "Auscultation of the respiratory system," *Ann Thorac Med*, vol. 10, no. 3, pp. 158–168, 2015, doi: 10.4103/1817-1737.160831.

[5] "Lung Sounds," Cleveland Clinic. Accessed: Oct. 25, 2024. [Online]. Available: https://my.clevelandclinic.org/health/symptoms/25193-lung-sounds

[6] M. T. Nguyen, W. W. Lin, and J. H. Huang, "Heart Sound Classification Using Deep Learning Techniques Based on Log-mel Spectrogram," *Circuits Syst Signal Process*, vol. 42, no. 1, pp. 344–360, Jan. 2023, doi: 10.1007/s00034-022-02124-1.

[7] M. Xiang *et al.*, "Research of heart sound classification using two-dimensional features," *Biomedical Signal Processing and Control*, vol. 79, p. 104190, Jan. 2023, doi: 10.1016/j.bspc.2022.104190.

[8] N. Faruqui, M. A. Yousuf, M. Whaiduzzaman, A. K. M. Azad, A. Barros, and M. A. Moni, "LungNet: A hybrid deep-CNN model for lung cancer diagnosis using CT and wearable sensor-based medical IoT data," *Computers in Biology and Medicine*, vol. 139, p. 104961, Dec. 2021, doi: 10.1016/j.compbiomed.2021.104961.

[9] "Sensors | Free Full-Text | Feature-Based Fusion Using CNN for Lung and Heart Sound Classification." Accessed: Aug. 14, 2024. [Online]. Available: https://www.mdpi.com/1424-8220/22/4/1521

[10] Y. Choi and H. Lee, "Interpretation of lung disease classification with light attention connected module," *Biomed Signal Process Control*, vol. 84, p. 104695, Jul. 2023, doi: 10.1016/j.bspc.2023.104695.

[11] "ICBHI 2017 Challenge | ICBHI Challenge." Accessed: Aug. 14, 2024. [Online]. Available: https://bhichallenge.med.auth.gr/ICBHI_2017_Challenge

[12] M. Fraiwan, L. Fraiwan, B. Khassawneh, and A. Ibnian, "A dataset of lung sounds recorded from the chest wall using an electronic stethoscope," *Data in Brief*, vol. 35, p. 106913, Apr. 2021, doi: 10.1016/j.dib.2021.106913.

[13] Y. Kim *et al.*, "Respiratory sound classification for crackles, wheezes, and rhonchi in the clinical field using deep learning," *Sci Rep*, vol. 11, p. 17186, Aug. 2021, doi: 10.1038/s41598-021-96724-7.

[14] W.-B. Ma, X.-Y. Deng, Y. Yang, and W.-C. Fang, "An Effective Lung Sound Classification System for Respiratory Disease Diagnosis Using DenseNet CNN Model with Sound Pre-processing Engine," in *2022 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, Oct. 2022, pp. 218–222. doi: 10.1109/BioCAS54905.2022.9948568.

[15] J. Acharya and A. Basu, "Deep Neural Network for Respiratory Sound Classification in Wearable Devices Enabled by Patient Specific Model Tuning," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 14, no. 3, pp. 535–544, Jun. 2020, doi: 10.1109/TBCAS.2020.2981172.

[16] S. B. Shuvo, S. N. Ali, S. I. Swapnil, T. Hasan, and M. I. H. Bhuiyan, "A Lightweight CNN Model for Detecting Respiratory Diseases From Lung Auscultation Sounds Using EMD-CWT-Based Hybrid Scalogram," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 7, pp. 2595–2603, Jul. 2021, doi: 10.1109/JBHI.2020.3048006.

[17] T. Nguyen and F. Pernkopf, "Lung Sound Classification Using Snapshot Ensemble of Convolutional Neural Networks," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Jul. 2020, pp. 760–763. doi: 10.1109/EMBC44109.2020.9176076.

[18] F. Wang, X. Yuan, and B. Meng, "Classification of Abnormal Lung Sounds Using Deep Learning," in *2023 8th International Conference on Signal and Image Processing (ICSIP)*, Jul. 2023, pp. 506–510. doi: 10.1109/ICSIP57908.2023.10271089.

[19] H. S. Kim and H. S. Park, "Ensemble Learning Model for Classification of Respiratory Anomalies," *J. Electr. Eng. Technol.*, vol. 18, no. 4, pp. 3201–3208, Jul. 2023, doi: 10.1007/s42835-023-01425-y.

[20] D. Bardou, K. Zhang, and S. M. Ahmad, "Lung sounds classification using convolutional neural networks," *Artificial Intelligence in Medicine*, vol. 88, pp. 58–69, Jun. 2018, doi: 10.1016/j.artmed.2018.04.008.

[21] Y.-S. Wu, C.-H. Liao, and S.-M. Yuan, "Automatic Auscultation Classification of Abnormal Lung Sounds in Critical Patients Through Deep Learning Models," in *2020 3rd IEEE International Conference on Knowledge Innovation and Invention (ICKII)*, Aug. 2020, pp. 9–11. doi: 10.1109/ICKII50300.2020.9318880.

[22] N. Zakaria and K. Sundaraj, "VGG16-Based Deep Learning Architectures for Classification of Lung Sounds into Normal, Crackles, and Wheezes using Gammatonegrams," in *2023 International Conference on Information Technology (ICIT)*, Aug. 2023, pp. 83–88. doi: 10.1109/ICIT58056.2023.10225790.

[23] P. Faustino, J. Oliveira, and M. Coimbra, "Crackle and wheeze detection in lung sound signals using convolutional neural networks," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Nov. 2021, pp. 345–348. doi: 10.1109/EMBC46164.2021.9630391.

[24] H. Chanane and M. Bahoura, "Convolutional Neural Network-based Model for Lung Sounds Classification," in *2021 IEEE International Midwest Symposium on Circuits and Systems (MWSCAS)*, Aug. 2021, pp. 555–558. doi: 10.1109/MWSCAS47672.2021.9531887.

Authorized licensed use limited to: University of Plymouth. Downloaded on November 24,2025 at 09:58:19 UTC from IEEE Xplore. Restrictions apply.