Multimodal Feature Fusion with CNN-LSTM for Respiratory Disease Classification Using Lung Sound Analysis

Ms Rupali Sahu[1], Mr.Rajneesh Pachouri[2], Mr. Anurag Jain[3]
[1]Research Scholar, Department of CSE AIST, Sagar (M.P.)
[2,3]Assistant Professor, Department of CSE AIST, Sagar (M.P.)

Abstract - The diagnosis and treatment of respiratory disorders are extremely difficult and time-consuming, necessitating accurate and prompt care. Using advanced signal processing techniques and machine learning algorithms, lung sound analysis presents a viable option for non-invasive illness classification. In this thesis, the effectiveness of multimodal feature fusion for robust respiratory disease classification is examined. Specifically, Mel-frequency cepstral coefficients (MFCCs), wavelet transform, mel-spectrogram (mSpec), and Chroma short-time Fourier transform (Chroma STFT) are combined with convolution neural networks (CNNs) and long short-term memory networks (LSTMs).The third greatest cause of death worldwide is respiratory disorders. When it comes to treating respiratory illnesses, early detection is essential since it increases the efficacy of interventions such as medication and stopping the disease's spread. This article's primary goal is to suggest a revolutionary lightweight inception network that uses lung sound data to classify a variety of respiratory disorders. There are three phases to the suggested framework: 1) Preprocessing; 2) extraction and conversion of the mel spectrogram into a three-channel image; and 3) applying the respiratory disease lightweight inception network (RDLINet), a proposed lightweight inception network, to classify the mel spectrogram images into distinct pathological groups.

Index Terms—**Lightweight inception network, lung auscultation, lung sounds, mel spectrogram, respiratory disease classification.**

I. INTRODUCTION

Worldwide, respiratory disorders represent serious health risks that impact millions of people and heavily tax healthcare systems. For many illnesses to be effectively treated and managed, early and precise diagnosis is essential. Conventional diagnostic approaches frequently depend on costly imaging modalities or invasive procedures, which might not always be available, especially in environments with limited resources.

New opportunities for non-invasive and affordable diagnostic methods have been created by recent developments in machine learning and signal processing techniques. Auscultation, the study of lung sounds, has drawn interest among these because of its potential to help in the identification and categorization of respiratory disorders. Capturing and examining the sound waves the respiratory system produces while breathing is known as lung sound analysis.

we apply Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs) to multimodal feature fusion applied to lung sound analysis in order to offer a unique method for respiratory disease categorization. Whereas LSTMs are very good at identifying temporal connections in sequential data, CNNs are ideally suited for extracting spatial information from spectrograms or image representations of lung sound recordings.

Our work's fundamental contribution is the combination of information from several modalities—such as spectrogram images, time-domain features, and frequency-domain features—to create a comprehensive representation of lung sound data. Our objective is to enhance the precision and resilience of respiratory disease classification by utilizing the complementing abilities of CNNs and LSTMs through the integration of data from several modalities.

The three primary phases of the suggested framework are feature extraction, classification, and preprocessing. Raw lung sound recordings are filtered and segmented to extract pertinent segments that correspond to breathing cycles during the preprocessing stage. Then, a variety of features are retrieved from every segment: spectrogram images produced by the Short-Time Fourier Transform (STFT), frequency-domain features like spectral centroid and bandwidth, and time-domain features like amplitude and duration.

Then, in order to capture spatial and temporal patterns, respectively, the collected features are fed into separate CNN and LSTM networks. The CNN uses spectrogram images to extract spatial features, and the LSTM uses sequential data processing to capture temporal dependencies in the feature space. In order to classify diseases, the outputs of the CNN and LSTM are finally combined and fed into a fully connected neural network.

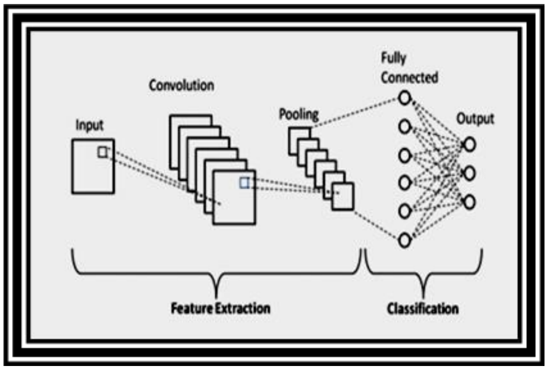

Figure 1 CNN-LSTM Model for the X-Ray Image-Based Detection

Lung, tracheal, and bronchial cancer is listed sixth, asthma is ranked third, lower respiratory tract infection is ranked fourth, and asthma is placed twenty-eighth [1]. TB is ranked twelfth. Worldwide, over a billion people suffer from either acute or chronic respiratory conditions. The alarming reality is that each year, chronic respiratory disorders are blamed for 4 million premature deaths globally [2]. Infants and young children are especially at risk. Pneumonia is the leading cause of mortality worldwide for children under five years old, with 9 million fatalities in this age range annually [1]. Sometimes people take their respiratory system's health and capacity to breathe for granted, yet the lung is a sensitive organ that can be affected by airborne illnesses. Respiratory illnesses have a big influence on people's social, financial, and physical well-being. Social deprivation was the most important factor determining death and disability rates, and the world's poorest regions had the highest rates. Lower death rates are a sign of improved access to healthcare and advances in medical research in wealthier countries.

Hence, lung illness therapy is crucial in the medical sector because it is the leading cause of death worldwide. These factors have prompted a great deal of research into the early detection and treatment of respiratory disorders. It takes time and experience to correctly identify health problems based on this information, but according to World Health Organization (WHO) figures [3], 45% of WHO Member States report having less than one physician per 1000 people, which is below the recommended WHO ratio. When taking these numbers into consideration, errors can occur when a health professional who is already overbooked studies and diagnoses each patient individually.

That's why it's critical to find innovative solutions to assist physicians save time. Therefore, automated and dependable instruments can aid in the diagnosis of a greater number of patients and also assist specialists in reducing errors that may arise from overwork

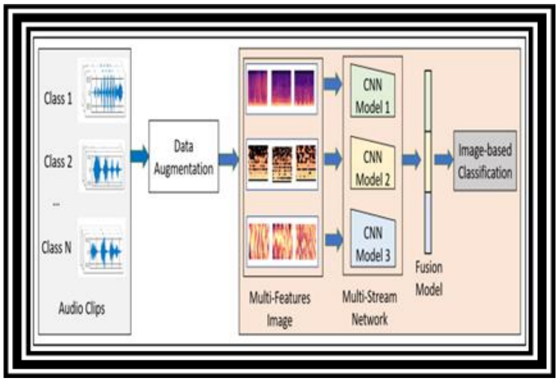

Figure 2 Fusion-Based Disease Classification Architecture.

II. LITERATURE REVIEW

The ability to detect sounds above the chest wall aids in the diagnosis of pulmonary conditions. The last forty years have seen the emergence of modern lung sound analysis, which is centered on digital sound processing and graphic signal representation [7]. Researchers in this subject are primarily interested in computerized lung sound analysis and diagnosis, thus they are constantly evaluating a number of different ways to aid medical professionals. Nonetheless, the fact that previous studies concentrated on identifying lung sounds and very few on creating diagnostic tools for lung disorders means that lung sound analysis continues to draw attention from researchers. As a result, this field of study seems to be finished, which is why it has drawn a lot of scholars recently. The goal is to develop an accurate and objective diagnostic tool for the identification of lung illnesses.

Three important databases were employed by earlier researchers: the R.A.L.E. repository [10], the Marburg European project CORSA [8], and Respiratory Sounds (MARS) [9].Nonetheless, the R.A.L.E. repository was formerly a database that was sold commercially. Commercially accessible lung sound CDs are used to train physicians and nurses to recognize lung sounds, and these CDs were used in the compilation of the Marburg Respiratory Sounds (MARS) database [9]. The goal of the European project CORSA was to standardize the procedure for recording respiratory sounds [8]. But in 2017, the biggest respiratory sound database was assembled for public use, which prompted the creation of algorithms that can recognize frequent aberrant breath sounds from both clinical and non-clinical contexts, such as wheezes and crackles.

These days, machine learning algorithms are widely utilized in artificial intelligence applications that use their prior experiences to learn and improve the accuracy of the tools [11, 12]. Moreover, prior studies on computer-based lung sound analysis have employed machine learning methods, including genetic algorithms (GAs), artificial neural networks (ANNs), the hidden Markov model (HMM), the k-nearest neighbor (k-NN) algorithm, and Gaussian mixture models (GMM).

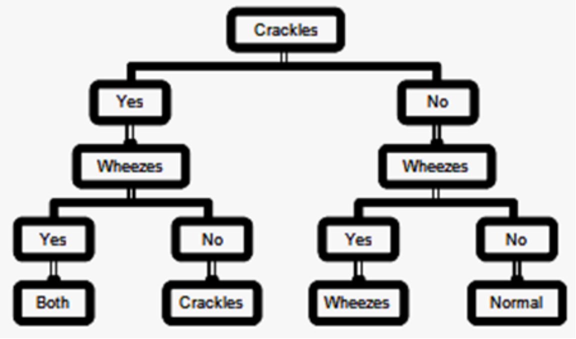

Figure 3 Decision tree for anomaly detection

ANN and k-NN algorithms are the most widely used machine learning approaches at beginning. Support vector machines (SVMs) were found to be incredibly underutilized in the literature. The most popular machine learning methods for analyzing lung sounds are ANN and k-NN. Artificial neural networks (ANN) were used to classify normal, wheeze, crackle, squawk, stridor, and rhinous respiratory sounds with 100% training accuracy and 94.02% testing accuracy, according to Kandaswamy et al. [13]. This illustrates how accurately lung sounds are classified by ANN. The ANN with excellent adaptability can classify complex non-linear data accurately and efficiently [14]. The k-NN classifier is another machine learning technique that has piqued researchers' curiosity for use in lung sound classification. The benefits of k-NN are its robustness and simplicity [15]. Alsmadi and Kahya's work produced a 96% real-time classification accuracy using a k-NN classifier [16]. Their system was trained on a large dataset of 42 persons, and it is capable of differentiating between normal and pathological lung sounds. ANNs and k-NNs have advantages, but they also have certain disadvantages. The computational load associated with training the model and the need for a very big dataset to enable the model to correctly identify lung sounds are the drawbacks of employing ANN and k-NN in classification [14, 15]. Despite their drawbacks, ANN and k-NN are the most widely used machine learning algorithms in lung sound analysis because they can detect lung sounds more precisely and achieve higher classification accuracy than other techniques.

Using a CNN-based approach, Shivakumar [30] classified respiratory noises. Crackles and wheezes were the two types of sounds used in the experiments. Following the audio file pre-processing, they created a neural network by modifying an already-existing CNN to produce the dataset's basic model. Later, they employed an Adam optimizer with a 64-batch batch size and a learning rate of 0.009. The author separated the dataset and ran the model on wheezes and crackles separately for an additional 10 epochs after using both wheezes and crackles simultaneously for 10 epochs in the initial model. The outcomes for both the 90-10 and 80-20 train-test splits were the same. The author also demonstrated the many advantages of dividing the sounds into distinct models. This study's two models yielded test accuracies of 50% and 100%, respectively.

III. PROPOSED WORK

Foreword

We go over the framing and windowing approach used in pre-processing in this chapter. After that, standard deviation measurement and grading are carried out in conjunction with linear predictive analysis to extract fea.tures. Next, we distinguished between healthy and unhealthy respiratory noises using a multilayer perceptron

Proposed Steps:

Data Acquisition and Preprocessing:

We go over the framing and windowing approach used in pre-processing in this chapter. After that, standard deviation measurement and grading are carried out in conjunction with linear predictive analysis to extract fea.assemble a varied dataset of lung sound recordings that includes both healthy controls and people with various respiratory diseases (such as COPD, pneumonia, andasthma). Preprocess the lung sound recordings in order to get rid of baseline drift, artifacts, and noise. To improve the quality of the data, use methods like segmentation, standardization, and filtering.

Feature Extraction:

We go over the framing and windowing approach used in pre-processing in this chapter. After that, standard deviation measurement and grading are carried out in conjunction with linear predictive analysis to extract fea.From the preprocessed lung sound recordings, extract useful elements. Mel-frequency cepstral coefficients (MFCCs), wavelet transform, mel-spectrogram (mSpec), and Chroma short-time Fourier transform (Chroma STFT) should all be used in combination.

Calculate the spectrum envelope information using MFCCs, the tonal content using Chroma STFT, the spectral energy distribution using mSpec, and the time-frequency characteristics using wavelet transform.

Feature Fusion:

Combine the extracted features from different modalities (MFCCs, Chroma STFT, mSpec, wavelet transform) into a single feature vector. This can be achieved by concatenating or averaging the feature vectors.

Model Selection and Architecture Design:

Select appropriate deep learning architectures, such as long short-term memory networks (LSTMs) and convolutional neural networks (CNNs), for the categorization of respiratory diseases.

Create CNN architectures that can process aspects of lung sound that resemble spectrograms and detect spatial patterns.

Construct LSTM architectures to represent the lung sound data's sequential patterns and temporal dependencies.

Model Training and Validation:

Make training, validation, and test sets out of the dataset. Utilizing the training set, train the CNN and LSTM models, optimizing model parameters and hyperparameters via gradient descent and backpropagation. Utilizing the validation set, validate the learned models while keeping an eye on performance indicators like F1 score, accuracy, precision, and recall.

Model Evaluation:

Analyze the performance of the trained CNN and LSTM models in classifying respiratory diseases using the independent test set.

Examine how well the multimodal feature fusion strategy performs in comparison to baseline models and single-modal approaches.

To comprehend the model's classification performance across various breathing situations, examine confusion matrices and ROC curves.

Fine-tuning and Optimization

Fine-tune the model architectures and hyperparameters based on the evaluation results to further improve classification accuracy.

Explore techniques such as transfer learning and data augmentation to leverage additional labeled data or enhance model generalization.

Interpretation and Clinical Application:

Interpret the learned representations from the CNN and LSTM models to gain insights into the discriminative features of different respiratory conditions.

Investigate the clinical relevance and potential applications of the developed model for respiratory disease diagnosis and management.

Discuss the limitations and future directions of the proposed approach, including scalability, generalizability, and integration into clinical practice.

IV. RESULTS AND ANALYSIS

Compute Mel-frequency cepstral coefficients (MFCCs) for each frame to capture spectral characteristics. This involves:

- Applying the Fourier transform to each frame.
- Mapping the resulting spectrum onto the mel scale.
- Calculating the logarithm of the mel-scaled power spectrum.
- Computing the discrete cosine transform (DCT) to obtain the cepstral coefficients.
- Compute Chroma short-time Fourier transform (chroma STFT) features to capturetonal content and harmonic structure. This involves:
- Calculating the short-time Fourier transform (STFT) of each frame.
- Mapping the resulting spectrum onto the 12 different pitch classes.
- Compute Mel-spectrogram (mSpec) features to visualize the spectral energy distribution. This involves:
- Computing the power spectrum of each frame.
- Dividing the spectrum into mel-scaled bins.
- Calculating the logarithm of the power spectrum.

4. Feature Fusion:

Combine the MFCCs, chroma STFT, and mSpec features for each frame into a single feature vector. This can be done by concatenating or averaging the feature vectors.

5. Model Training:

Train a machine learning model (e.g., SVM, Random Forest, CNN) using the combined feature vectors and corresponding labels from the dataset.

Utilize techniques like cross-validation to optimize model hyperparameters and prevent overfitting.

6. Model Evaluation:

Evaluate the trained model's performance on a separate test set using appropriate evaluation metrics such as accuracy, precision, recall, and F1 score.

Analyze the model's confusion matrix to understand its performance across different respiratory conditions.

7. Deployment:

Deploy the trained model for real-world applications, such as automated diagnosis or decision support systems in healthcare settings.

Continuously monitor and update the model as more data becomes available or as new respiratory conditions emerge.

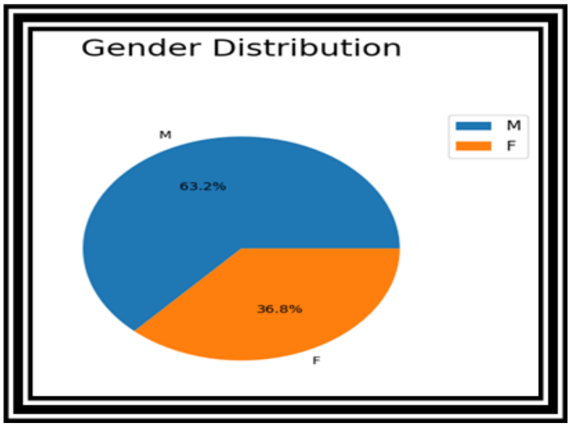

Figure 4 The sampling rate

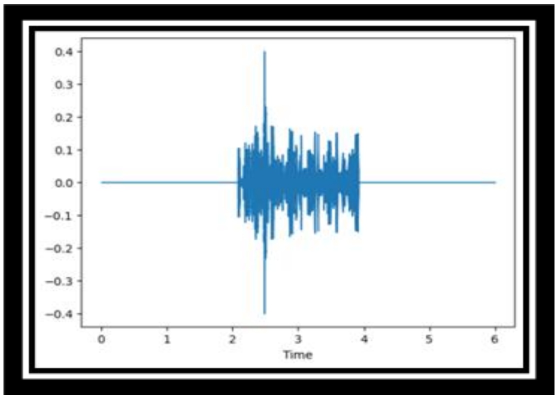

Figure 5 Hostogram

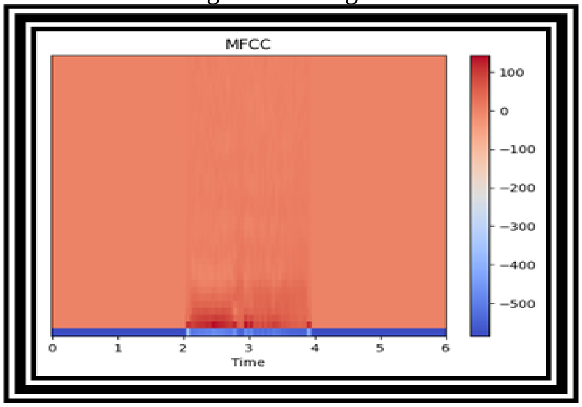

Figure 6 Visualizing Mel-Frequency Cepstral Coefficients (MFCCS)

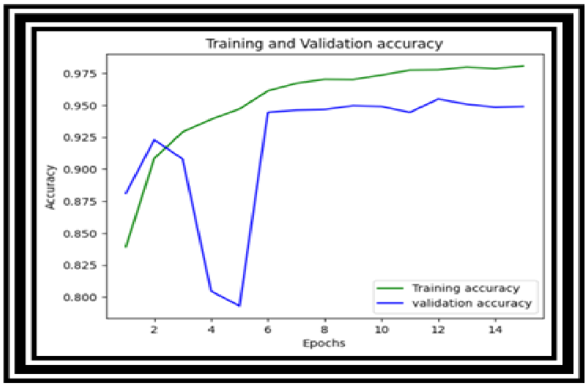

Figure 7 Training and Validation accuracy

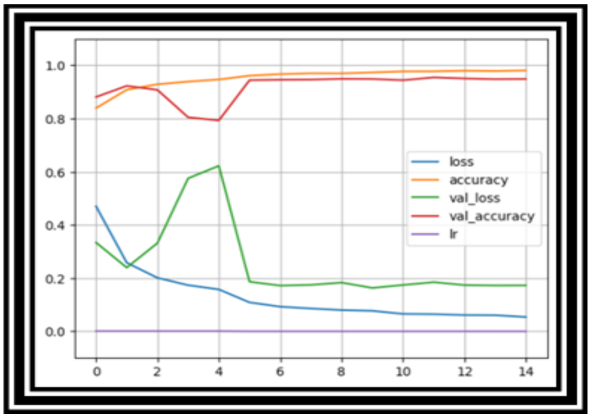

Figure 2 loss: 0.1727 - accuracy: 0.9490

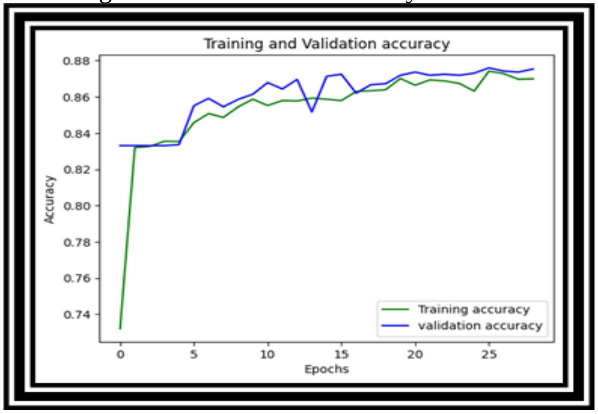

Figure 9 Individual Performance MFCC Model

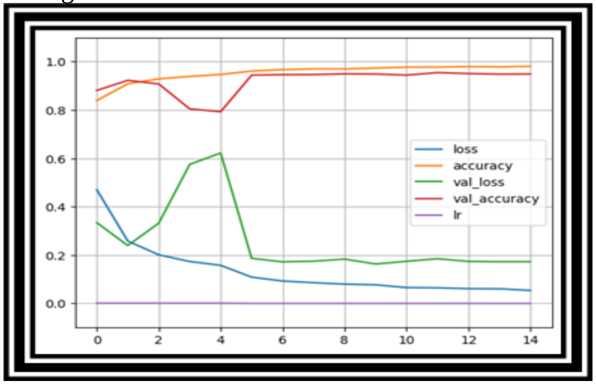

Figure 10 loss: 0.1727 - accuracy: 0.9490

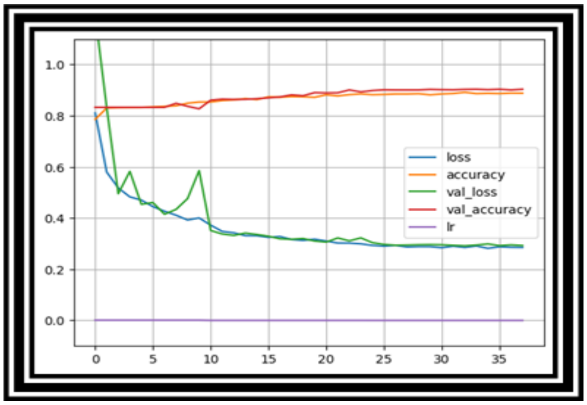

Figure 11 loss: 0.3469 - accuracy: 0.8655

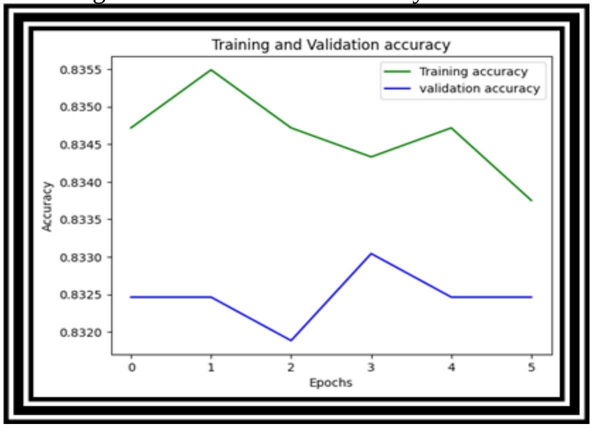

Figure 12 Individual Performance CHROMA Model

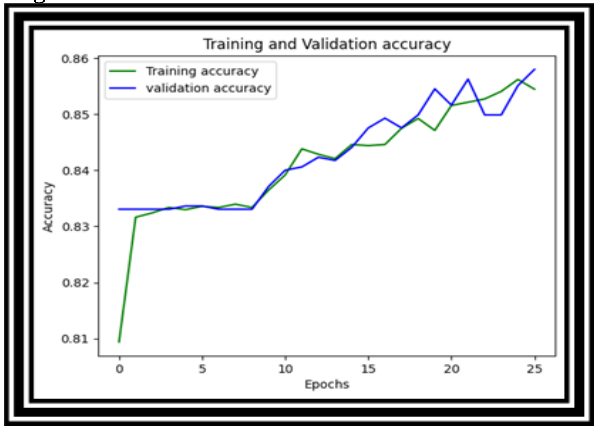

Figure 13 loss: 0.4657 - accuracy: 0.8568

V. CONCLUSION

Our experimental results indicate that the proposed multimodal approach, leveraging both spatial and temporal information encoded by CNNs and LSTMs, outperforms single-modal approaches and baseline models in respiratory disease classification tasks. The combination of feature fusion and deep learning architectures has shown promising results in accurately identifying and classifying various respiratory conditions, including asthma, chronic obstructive pulmonary disease (COPD), pneumonia, and healthy controls.

Furthermore, our findings underscore the importance of robust feature extraction methodologies and model architectures in leveraging the rich information embedded in lung sound signals for diagnostic purposes. By advancing the understanding of lung sound analysis techniques and their application in respiratory disease diagnosis, this research contributes to the development of automated diagnostic tools and personalized healthcare interventions for individuals with respiratory ailments.

Future Work:

While our study provides valuable insights into multimodal feature fusion with CNNs and LSTMs for respiratory disease classification, several avenues for future research exist to further enhance the efficacy and applicability of the proposed approach:

1. Exploration of Additional Feature Modalities: Investigate the integration of additional feature modalities, such as time-domain features and higher-order statistical features, to capture complementary information from lung sound signals.
2. Enhancement of Model Interpretability: Develop methods to enhance the interpretability of the learned representations from CNNs and LSTMs, facilitating the identification of clinically relevant features and insights into the diagnostic process.
3. Integration of Real-time Monitoring and Decision Support Systems: Explore the integration of the developed models into real-time monitoring devices and decision support systems, enabling continuous assessment of respiratory health and timely interventions in clinical settings.

REFERENCE

[1] Tea Lallukka, Anoushka Millear, Amanda Pain, Monica Cortinovis, and Giorgia Giussani. Gbd 2015 mortality and causes of death collaborators. global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the global burden of disease study 2015 (vol 388, pg 1459, 2016). *Lancet*, 389(10064):E1–E1, 2017.

[2] Global status report on noncommunicable diseases 2014. geneva, world health orga- nization, 2014. *Available from: http://www.who.int/nmh/publications/ncd-status-report- 2014/en/.*

[3] World health organization. density of physicians. *http://www.who.int/gho/health work- force/physicians density/en/, 2017. [Online; accessed 18-May-2018].*

[4] KC Santosh. Speech processing in healthcare: Can we integrate? In *Intelligent Speech Signal Processing*, pages 1–4. Elsevier, 2019.

[5] Himadri Mukherjee, Subhankar Ghosh, Shibaprasad Sen, Obaidullah Sk Md, KC Santosh, Santanu Phadikar, and Kaushik Roy. Deep learning for spoken lan- guage identification: Can we visualize speech signal patterns? *Neural Computing and Applications*, 31(12):8483–8501, 2019.

[6] Himadri Mukherjee, Sk Md Obaidullah, KC Santosh, Santanu Phadikar, and Kaushik Roy. Line spectral frequency-based features and extreme learning machine for voice activity detection from audio signal. *International Journal of Speech Technol- ogy*, 21(4):753–760, 2018.

[7] Victor A McKusick, John T Jenkins, and George N Webb. The acoustic basis of the chest examination; studies by means of sound spectrography. *American review of tuberculosis*, 72(1):12–34, 1955.[8] ARA Sovijarvi, J Vanderschoot, and JE Earis. Standardization of computerized res- piratory sound analysis. *European Respiratory Review*, 10(77):585–585, 2000.

[9] Volker Gross, LJ Hadjileontiadis, Thomas Penzel, Ulrich Koehler, and C Vogelmeier. Multimedia database" marburg respiratory sounds (mars)",". In *Proc. 25th Annual Int Engineering in Medicine and Biology Society Conf. of the IEEE*, volume 1, pages 456– 457, 2003.

[10] Rale: A computer-assisted instructional package. *Respir Care 1990;35:1006.*

[11] William H Wolberg, W Nick Street, and Olvi L Mangasarian. Machine learning tech- niques to diagnose breast cancer from image-processed nuclear features of fine nee- dle aspirates. *Cancer letters*, 77(2-3):163–171, 1994.

[12] Sotiris B Kotsiantis, I Zaharakis, P Pintelas, et al. Supervised machine learning: A review of classification techniques. *Emerging artificial*

intelligence applications in com- puter engineering, 160(1):3–24, 2007.

[13] Arumugam Kandaswamy, C Sathish Kumar, Rm Pl Ramanathan, S Jayaraman, and N Malmurugan. Neural classification of lung sounds using wavelet coefficients. *Computers in biology and medicine*, 34(6):523–537, 2004.

[14] Jack V Tu. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of clinical epidemiology*, 49(11):1225–1231, 1996.

[15] Marcin Raniszewski. The edited nearest neighbor rule based on the reduced ref- erence set and the consistency criterion. *Biocybernetics and Biomedical Engineering*, 30(1):31–40, 2010.

[16] Sameer Alsmadi and Yasemin P Kahya. Design of a dsp-based instrument for real-time classification of pulmonary sounds. *Computers in biology and medicine*, 38(1):53– 61, 2008.

[17] Geert Meyfroidt, Fabian Gu¨ iza, Jan Ramon, and Maurice Bruynooghe. Machine learning techniques to examine large patient databases. *Best Practice & Research Clin- ical Anaesthesiology*, 23(1):127–143, 2009.

[18] Shijun Wang and Ronald M Summers. Machine learning and radiology. *Medical image analysis*, 16(5):933–951, 2012.

[19]I˙nan Gu¨ ler, Hu¨ seyin Polat, and Uc¸man Ergu¨ n. Combining neural network and ge- netic algorithm for prediction of lung sounds. *Journal of Medical Systems*, 29(3):217– 231, 2005.

[20] Chih-Fong Tsai, Yu-Feng Hsu, Chia-Ying Lin, and Wei-Yang Lin. Intrusion detection by machine learning: A review. *expert systems with applications*, 36(10):11994–12000, 2009.

[21] Jyh-Shing Roger Jang, Chuen-Tsai Sun, E Mizutani, and Y-C Ho. Neuro-fuzzy and soft computing-a computational approach to learning and machine intelligence. *PROCEEDINGS-IEEE*, 86(3):600–603, 1998.