Journal of Information Technology Management
ISSN-OnLine: 2980-7972
University of Tehran

Attentional Deep Learning with Inverse Transform Sampling for Robust Respiratory Sound Classification

Hemanth K S *

*Corresponding author, Department of Computer Science, Christ University, Bengaluru, India. E-mail: hemanth.ks@christuniversity.in

Harisha Naik T

Department of Computer Applications, Presidency College, Bangalore, India. E-mail: harishtkola@gmail.com

N Kartik

Department of Commerce, Manipal Academy of Higher Education, Manipal, India. E-mail: kartik.n@manipal.edu

N.Nanda kumar

Assistant Professor, Department of Computer Science and Engineering, Excel Engineering College Komarpalayampalyam, Tamilnadu, India. E-mail: kumarnanda0079@gmail.com

S.Senthilkumar

Associate Professor, Department of Computer Science and Engineering, Vinayaka Mission`s Kirupananda Variyar Engineering College, Salem (Vinayaka Mission`s Research Foundation), India. E-mail: senthilkumars@vmkvec.edu.in

Ramya R

Associate professor, Department of Computer Science and Engineering, Bannari Amman Institute of Technology, Erode, Tamilnadu, India. E-mail: ramyarv@bitsathy.ac.in

Journal of Information Technology Management, 2025, Vol. 18, Issue 1, pp. 123-140
Published by the University of Tehran, College of Management
doi: https://doi.org/10.22059/jitm.2026.106257
Article Type: Research Paper

Received: August 08, 2025
Received in revised form: September 27, 2025
Accepted: December 14, 2025
Published online: January 20, 2026

Abstract

The necessity for efficient breathing sound classification systems originates from respiratory diseases, which impair oxygen-carbon dioxide exchange and impact lung function. Feature extraction and pattern categorization are general components of such systems. Because of their effectiveness with big datasets, deep neural networks have acquired popularity recently in the category of breathing sounds. Enhancing medical care requires cooperation amongst researchers, medical professionals, and patients. An attentional deep learning model with inverse transform sampling is presented in this study to classify respiratory diseases from audio data. Robust models were developed to classify and detect respiratory elements using the Respiratory Sound dataset. The primary objectives include effectively determining lung sounds and determining respiratory illnesses. The architectures of CNN, VGG16, and ResNet50 were developed to extract features and categorize data. Also, the pre-trained models ResNet50 and VGG16 identify critical characteristics in spectrum pictures more accurately. Inverse transfer sampling is used to rectify class imbalance in respiratory datasets. The models achieved 98% accuracy with the CNN model, 83% accuracy with VGG16, and 95% accuracy with ResNet50. Moreover, LSTM and CRNN models offer more information on how respiratory illnesses are classified.

Keywords: Respiratory Diseases, Deep Neural Network, Inverse Transform Sampling, CNN, Pre-trained Models.

Introduction

Early detection and treatment of lung disease are crucial for reducing sickness and mortality rates (Khanaghavalle et al., 2023). Lung sound provides valuable information about a patient's pulmonary status. Physicians utilize an auscultation technique to analyze lung sounds and determine the type of illness affecting an organ (Shuvo et al., 2020). The ailments include pneumonia, asthma, and bronchiectasis. Auscultation is a non-invasive therapy that helps diagnose diseases faster and improves treatment efficiency (Phettom et al., 2023). However, the intricate nature sound characteristic might lead to loss of data and incorrect diagnosis.

Lung sounds contain inherently non-stationary signals; thus, the existing method may not always provide an efficient response while acquiring sound. The stethoscope instrument used for auscultation is a sound channel that connects the human body's surface to the ear (Ma et al., 2024). Identifying sounds can be challenging owing to noise, ear sensitivity, and other factors that can impact diagnosis results. An automated system for identifying illnesses using lung sounds reduces subjectivity and provides an effective clinical diagnosis tool.

Respiratory noises are generated by air passing through the respiratory system during the inhaling and exhaling process. These breathing sounds can be captured from the thorax, trachea, or mouth (Tsai et al., 2020). Adventitious respiratory sounds (ARS) are aberrant respiratory noises that have been overlaid on normal respiratory sounds. They are divided into two types: continuous ARS (wheezes lasting more than 100 milliseconds) and discontinuous ARS (crackles lasting less than 20 milliseconds). ARS may help diagnose diseases based on the location, length, and strength of respiratory sounds (Phettom et al., 2023).

Over the last decade, various research methodologies have been presented and verified for the automated detection of respiratory problems using respiratory auscultation sounds. Several methods for obtaining feature data, which include statistical features (Tsai et al., 2020), entropy-based features (Koshta et al., 2021), wavelet coefficients, the Cepstral Coefficients (MFCC) (Phettom et al., 2023), spectrograms (Wang et al., 2023), and scalograms (Kilic et al., 2024), have been used in combination with various machine learning (ML) algorithms (Shuvo et al., 2020; Kili et al., 2024).

Recent improvements in deep learning (DL) have shown promising outcomes in clinical settings (Shuvo et al., 2020; Ali et al., 2023). Deep learning may overcome the constraints of classic ML-based techniques by utilizing autonomous feature learning. In recent years, DL-based techniques for diagnosing respiratory disorders and illnesses utilizing lung auscultation data have yielded promising results (Kilic et al 2024; Phettom et al., 2023). To perform optimally, deep networks require extensive training on large datasets. This demands a large amount of time and computer resources.

In this paper, CNN VGG16, ResNet50 CRNN, and LSTML are proposed to identify respiratory disorders using the respiratory sound database (Ma, W.B, et al., 2023), including patient-independent dataset splits for training, validation, and testing. A hybrid approach for creating scalograms from respiratory sound signals is described, in which CWT is performed exclusively on the best correlated intrinsic mode function (IMF) derived from the respiratory sound signals' EMD. This study created a hybrid neural model that employs the focal loss (FL) function to address class imbalance. The purpose of this study is to create a self-organizing, computationally efficient handmade model for automated asthma identification. Using the respiratory sound database, our model did well in binary categorization of the disease.

Literature Review

Previous research has extensively used machine learning and deep learning to classify respiratory sounds automatically. Most techniques have focused on predicting respiratory anomalies, such as wheezes and crackles (Khanaghavalle et al., 2023; Phettom et al., 2023; Kilic et al., 2024), rather than predicting respiratory illnesses directly from lung auscultation recordings. Recent techniques for disease classification rely on complex signal processing or specific CNN and RNN networks (Wang et al., 2023). Pathology-level characterization has been investigated at three levels: binary (healthy/pathological; Ma et al., 2024; Tsai et al., 2024), three-class chronic (healthy/chronic/non-chronic; Khanaghavalle et al., 2023), and multi-class different disease classification (Ali et al., 2023). Non-chronic diseases include Upper and Lower Respiratory Tract Infections (URTI and LRTI), bronchiolitis, and pneumonia, whereas chronic diseases include COPD, asthma, and bronchiectasis (Tsai et al., 2024).

Previous research has focused on identifying disorders from lung sounds and eliminating noise inter-ference, particularly from heart sounds (Phettom et al., 2023; Kilic et al., 2024). Research has con-centrated on noise removal approaches, especially for circumstances when heart sounds interact with lung sounds (Kilic et al., 2024; Tsai et al., 2024). These investigations have helped in creating noise-removal strategies and accurate illness detection from lung sounds to address issues of overlapping heart sounds and ambient noise. Several studies have been implemented for different approaches and algorithms to distinguish between normal and abnormal lung sounds, such as the Hough transform of spectrograms (Ali et al., 2023), wavelet packet decomposition (Koshta et al., 2021), adaptive com-plex in-exhale segmentation (Koshta et al., 2021), and time-expanded waveform analysis (Tariq et al., 2019). Advancements have enabled precise and automated evaluation of respiratory sounds, re-sulting in better detection and treatment of respiratory illnesses.

Methodology

Data Set

A large dataset is necessary for accurate lung sound categorization. This dataset should include lung sounds from as many distinct individuals as possible, regardless of age, gender, or respiratory health. The ICBHI 2017 Challenge (Khanaghavalle et al., 2023) was held during the 2017 International Symposium on Biomedical and Health Informatics. It was a technical competition with a breathing sound database and a defined scoring scheme. This collection includes annotated respiratory cycles from 126 people, totaling 5.5 hours of recordings. To simplify terminology, A cycle represents a patient's breathing cycle, where a record is an accumulation of lung sounds from an identical patient. The 126 different sounds were arranged into 6898 breathing cycles. Respiratory cycles varied from 0.2 to 16 seconds in duration. Table I shows the number of cycles observed.

Preprocessing

Preprocessing is an essential step preceding feature extraction and categorization. This study used the Band Pass Filter (Butterworth filter) to remove noise from lung sound signals (Wang et al., 2023). The initial stage of preparation involved reading audio recordings in.wav the format. The recorded sound files are imported into the library written in Python "librosa" and transformed into computerised signals. Because lung noises are often polluted in the real world, this method employs basic noise reduction based on a band-pass filter to preserve frequency band information important to lung sounds. It also suggests noise reduction techniques such as EMD, wavelet denoising, ICA, and so on. Segmentation. This step divides the recorded sound into intervals to create consistent data for training the extensive model. Each audio file has a respiratory cycle labelled, with problematic sounds from the lungs (crackles and wheezes) as 1 and others as 0. This module splits the recordings based on such labels. If the segment's duration is insufficient, wise padding (Anupama et al., 2024) or absolutely no padding is applied.

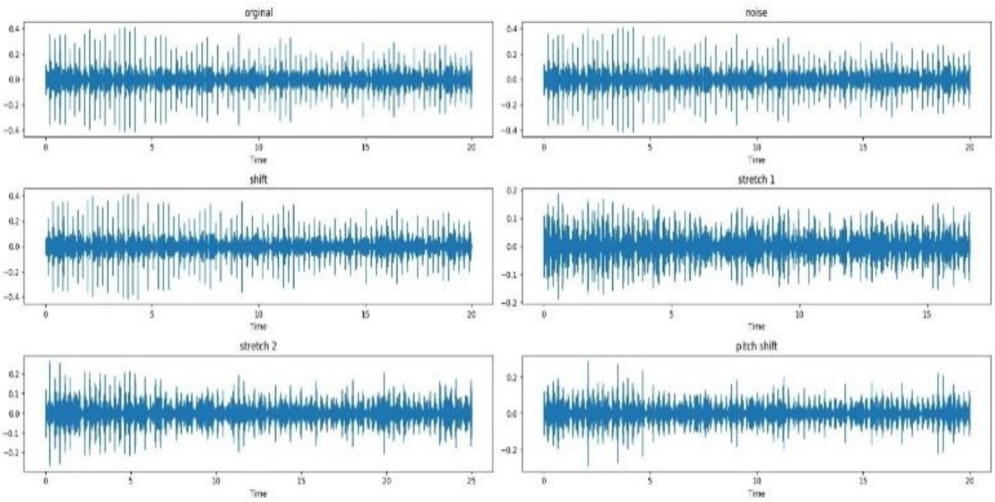

Figure 1. Audio files with various stretches

Feature extraction

Many variables contribute to the wide range of lung sounds, including age, gender, lung illness, and body positioning. The feature extraction approach is critical for producing Unique descriptions of features for categorization. As illustrated in Figure 2, lung sound representations use two methods of feature extraction: classic produced extraction and deep learning-based extraction (Koshta et al., 2023).

Traditional handcrafted traits comprise measurable audio signal characteristics that may be used to distinguish different sounds, which can be grouped into the following categories: (1) the time zone Features that identify information about Variations in lung sound over time include zero-crossing rate, root mean square, and wave surround; and (2) Harmonic-domain parameters such as spectral centroid, spectral roll-off, and spectral flux provide information regarding energy distribution over many frequency bands.

Mel-frequency cepstral coefficients (MFCCs) are an increasingly common characteristic in respiratory sound detection. obtained from the Fourier transform, which may identify Energy distribution across several ranges of frequency (Tariq et al., 2020) (Pham et al., 2023); and (3) time-frequency domain characteristics, which record its Energy distribution across different bands of frequencies throughout time, delivering valuable insights into the dynamic and lung sounds are fleeting in nature, including wavelet transform and spectrogram. It prioritises low-frequency components (less than 1 KHz) over high-frequency components (more than 1 KHz). The Mel scale represents frequency as

$$mel(f) \ = \ 2595. log\ (1 + 0) \quad (1)$$

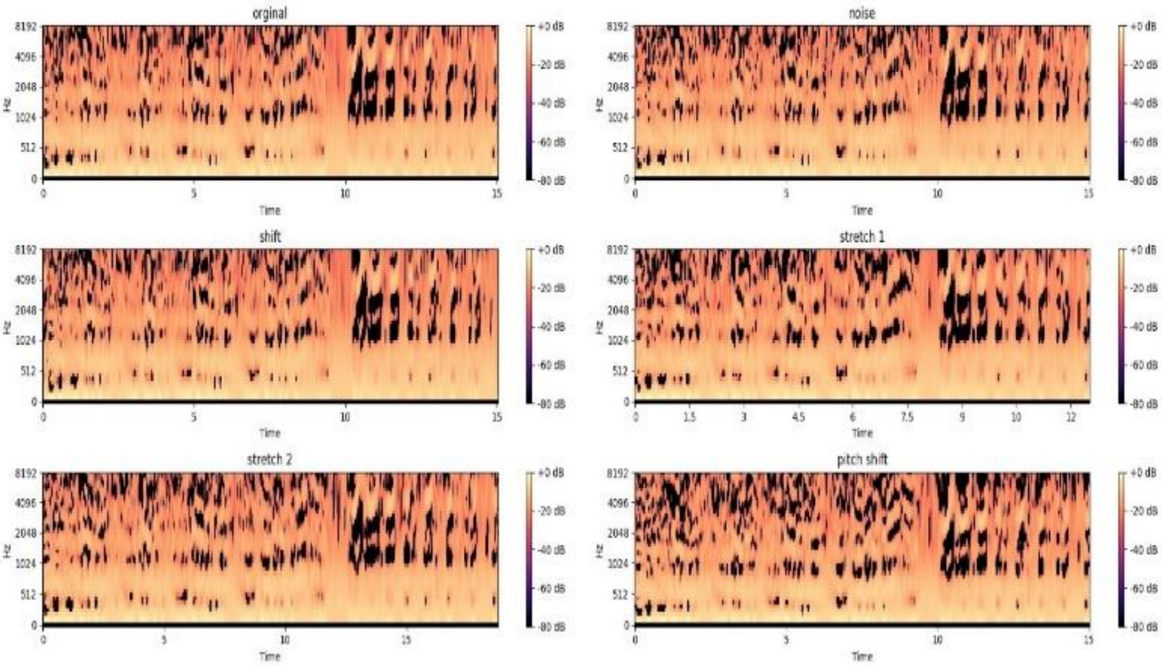

Figure 2. Extracting features from audio files represented with stretching.

There are several processes to calculate the MFCC. First, framing is utilized to deal with non-stationary signals like lung sounds, which are deemed stationary for short time periods. Frames are created by segmenting the lung signal using sliding windows. Additional research is needed to identify the optimal frame size and shift interval. In traditional audio transmissions, each frame consists of 256 samples, with an overlap of 128 samples. Changing the frame size and overlap between frames can assist in boosting performance. For example, a window size of 100ms with a 20ms overlap is common, while a window size of 0.5s with a 0.1s overlap is also effective. Additionally, the Hamming window is applied to each frame obtained during framing. The Hamming Window is described.

$$H(n) = 0.54 - 0.46{\sim}cos\ (\{2\backslash\pi \times n\}\ \{N - 1\}) \quad (2)$$

To enhance the speed of filtering, the Fast Fourier Transform (FFT) is employed to convert lung sounds from the time domain to the frequency domain. Creating a triangular window bank involves distributing a uniform triangular filter bank across the Mel-Warped spectrum. The Mel-spaced filter bank is generated based on the following relationships:

$$\Delta f\_\{mel\} = \{mel(f\_\{max\})\}/\{L\} \quad (3)$$

$$c(l) = 700[\ 10^\wedge l\Delta\, mel/2595 - 1] \quad (4)$$

The central frequency of the triangle filter is denoted by c(l) in Hertz in the equations above. The output energy of each filter is determined as follows:

$$m(l) = \sum\_\{k = o(l)\}\ ^\wedge\{k(l)\}\, W\_\{l\}(k)\ |X(k)| \quad (5)$$

In this study, instead of the Mel filter bank, we use a bespoke triangle filter with characteristics such as filter order, low frequency value, and high frequency value for the third triangle filter. To compute the MFCC, the logarithmic energy of the output is transformed using the Discrete Cosine Transformation (DCT).

Proposed Models

CNN model:

Convolution is an operation between two functions or signals. It is denoted with an asterisk and its one-dimensional is written as:

$$X\ s(t) = (x * w)\ (t) = a = -\infty\ x(a)w\ (t - a) \quad (6)$$

In this study of a convolutional neural network (CNN), the first signal, denoted as x, serves as the input, while the second signal, represented by w, corresponds to the filter. In this one-dimensional scenario, t represents the time index, and a signifies the time shift value. The resulting output of the convolutional layer is referred to as the feature map. In the case of images, it is more common to employ a two-dimensional convolution, which takes a three-dimensional matrix (comprising width, height, and color channels) as input and produces a corresponding three-dimensional matrix as output. The convolutional layer's parameters are a sequence of trainable filters that are convolved over the input's width and height, resulting in a two-dimensional feature map. The network will acquire filters that become activated upon detecting specific features.

The mathematical rationale behind employing convolutional neural networks (CNNs) for audio files is well-founded. When dealing directly with audio data, the input typically consists of a one-dimensional signal representing the waveform's amplitude over time. This one-dimensional signal is denoted as x(t), with (t) representing time.

One of the primary tasks of CNNs is the convolution process, where filters, also known as kernels, are applied in audio processing. Considering that k is typically smaller than w, let's designate F as the filter with a size of k. The convolution operation is outlined as follows:

$$(x * F) = \sum\_(a = 0)^\wedge(k - 1)\ \ x(t - a).F(a) \quad (7)$$

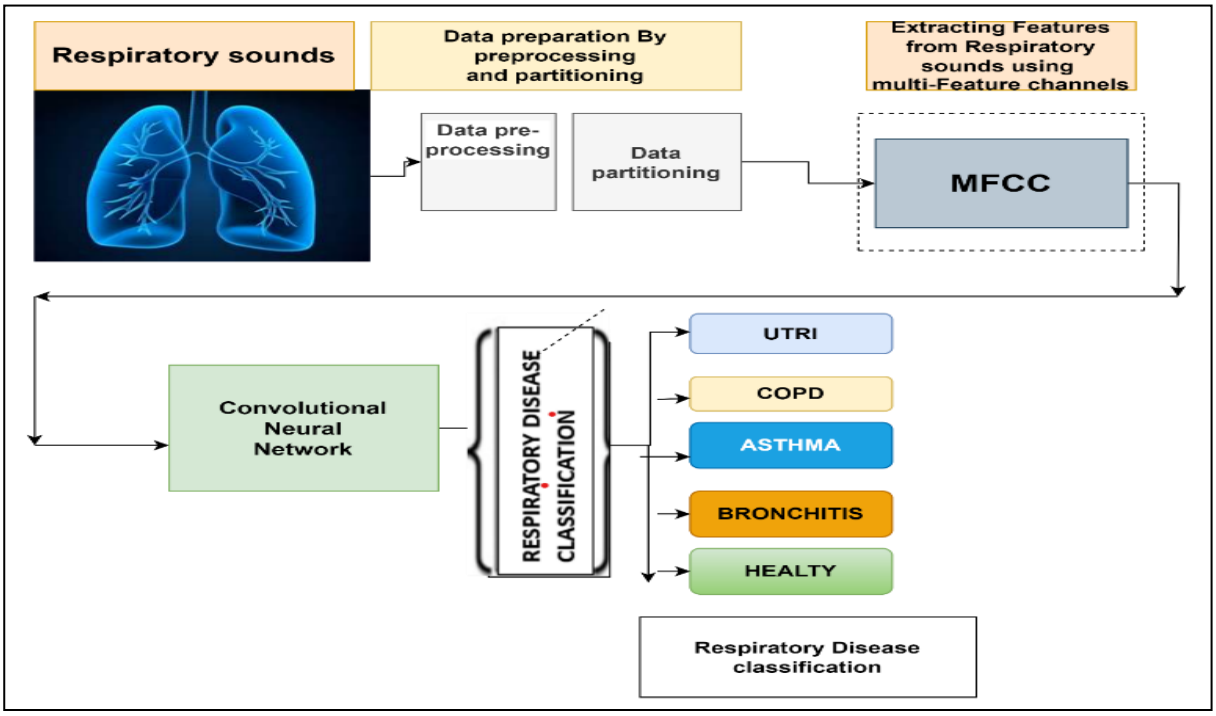

Figure 3. CNN designed for classifying respiratory lung sounds

Vgg16 model:

Convolutional neural networks, often known as ConvNets, are types of artificial neural networks that are artificial. A convolutional neural network consists of a training layer, an output layer, and several hidden layers. VGG16 is a kind of CNN (Convolutional Neural Network) that is regarded as one of the most powerful computer vision models to date. The architecture incorporates 3 x 3 filters, resulting in a substantial improvement over the previous layout. They increased the depth to 16-19 weight layers, yielding around 138 trainable parameters.

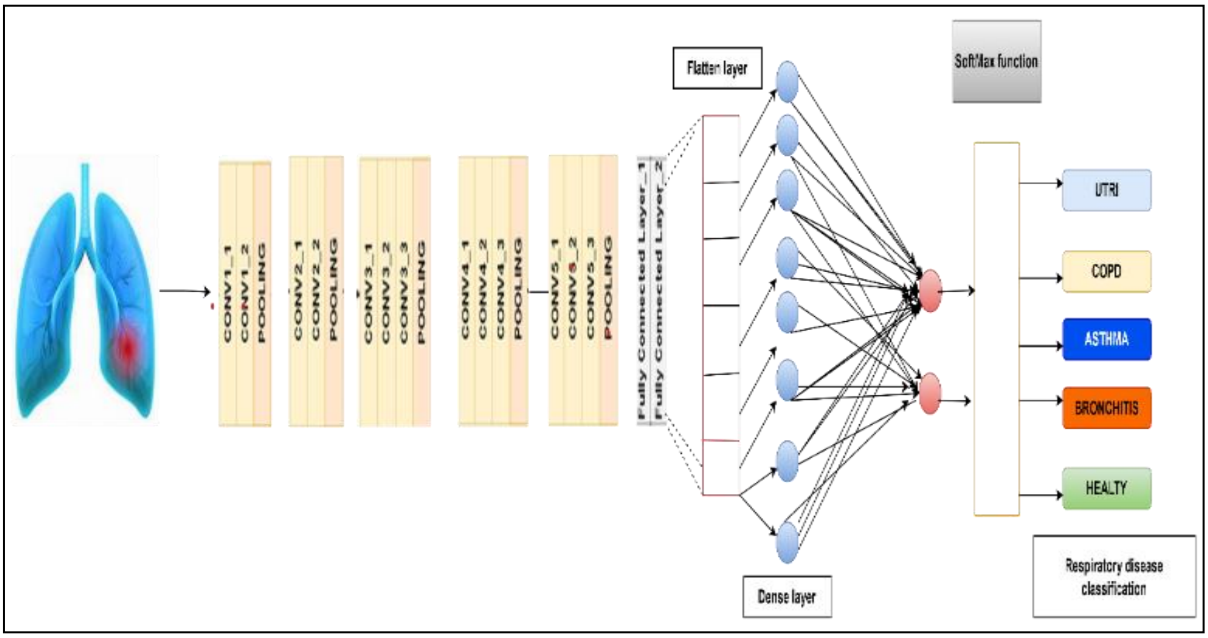

Figure 4. Diagram illustrating the architecture of VGG16 adapted for respiratory lung sound classification.

ResNet50

ResNet50 is a very effective image classification algorithm that can be trained on enormous datasets to produce cutting-edge results. Among its most significant advances is the use of residual connections, enabling the network to learn a set of latent functions for transforming input into desired output. These surviving connections allow the network to learn far deeper structures than previously feasible, while avoiding the problem of vanishing gradients.

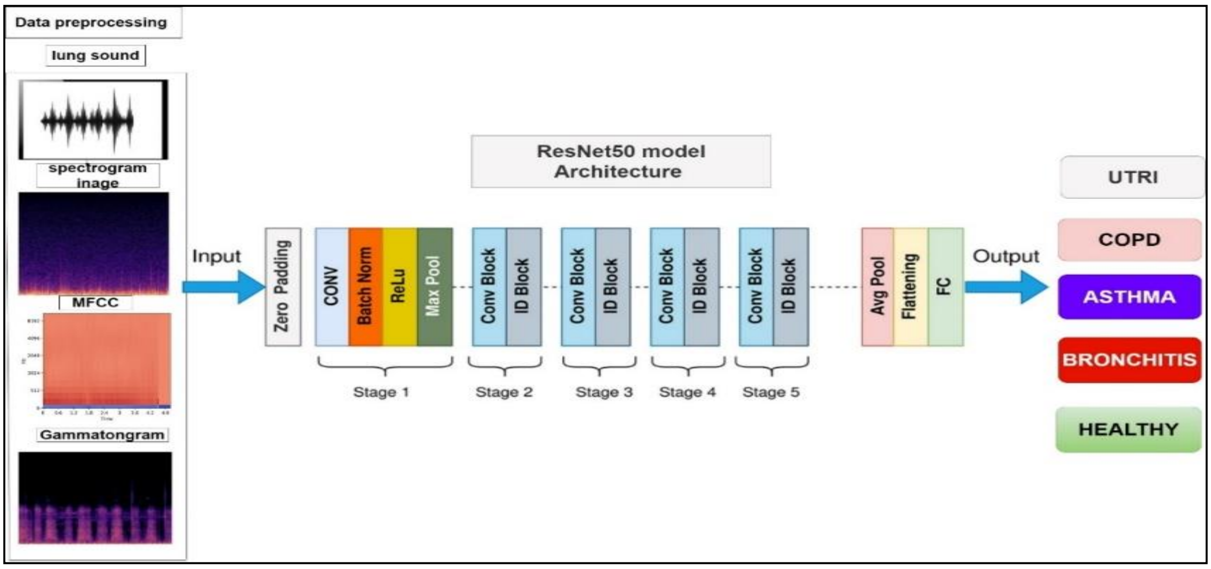

Figure 5. Diagram presenting the architecture of ResNet50 customized for respiratory lung sound classification

Input:

Let x be the input to the layer. For an audio file, x could represent a sequence of audio samples.

Linear Transformation (Fully Connected Layer):

$$z = Wx + b \tag{8}$$

W is the weight matrix.

b is the bias vector.

Activation Function

$$a = max\ (0, z) \tag{9}$$

a is the output after applying the activation function.

Normalization (Batch Normalization):

$$(a)^\wedge\ =\ (a - u)/\sigma \tag{10}$$

(a)ˆis the normalized output.

μ - mean

σ - standard deviation

Output:

The output y is then passed to the next layer.

CRNN:

Crnn Sound Classification transforms the input audio into a spectrum using a Mel spectrogram, then computes characteristics using CNN and LSTM, and then classifies it using FC and SoftMax.

The input parameter size is specified as (1, 2, number of seconds * sample rate). As an instance, given a 2-transmit wav data with 66026 samples (66026, 2) as data input, Audio Inference. Infer will convert it to (1,66026,1) using components (batch, samples, channel). This is then transformed to a Mel radio spectrum, and the output (1, 1, 128, 65), together with its different elements (batch, padding, bins, and times), serves as the source of data for the CNN. Using a similar method, the wav file (99225, 2) will be transformed to (1,1,128,97) as a feed to Conv.

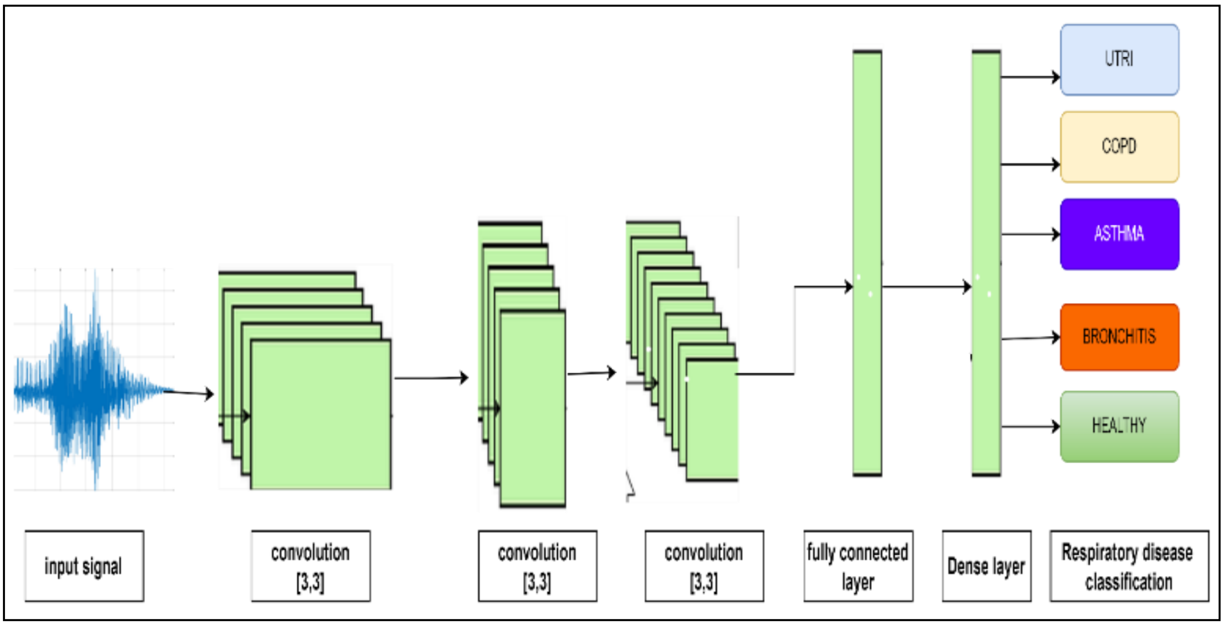

Figure 6. Diagram illustrating the architecture of a CRNN (Convolutional Recurrent Neural Network) tailored for respiratory lung sound classification.

LSTM:

LSTM, a key architectural design in artificial neural networks (ANNs), finds extensive use in deep learning (DL) and artificial intelligence (AI) domains. Notably, LSTM integrates feedback connections. Current research introduces a hybrid 2D CNN-LSTM network that integrates the FL function to address data imbalance. This network takes respiratory cycles as input and categorizes them into five distinct classes. Empowering it to proficiently process sequential data like time series and sound signals. Unique to LSTM are its specialized components known as "LSTM units", which encompass the input gate (it) is mainly for the getting the data, output gate (Ot) is for final layer, and forget gate is for remove the unwanted data (%&). Moreover, LSTM units employ the sigmoid function σ.

$$\Sigma = 1/1 + e^{\wedge}(-x) \tag{11}$$

Moreover, for the LSTM unit to operate effectively, it necessitates cells such as the cell state Ct, candidate state dt, and final output. Additionally, the procedure for the tanh function is given as well.

$$tanh = ((e^{\wedge}x - e^{\wedge}x))/((e^{\wedge}x + e^{\wedge}x)) \tag{12}$$

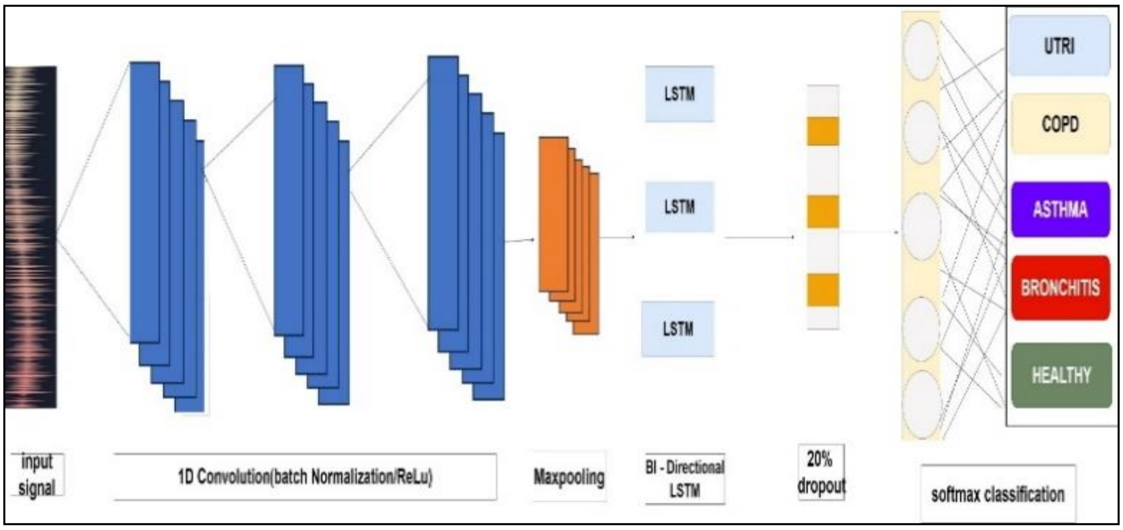

Figure 7. Diagram depicting the architecture of an LSTM (Long Short-Term Memory) network designed for respiratory lung sound classification.

Results

This work's results show that lung sounds can be detected. The accuracy of lung sound classification is determined by the output structure and the efficacy of the algorithm or model utilized.

Model performance was evaluated using standard metrics including accuracy, recall, Precision and F1-score geometric mean. The mathematical expressions for performance evaluation measures are shown below.

$$Ac = TN + TP / FP + TN + TP + FN \tag{13}$$

$$F1 = 2TP / 2TP + FP + FN \tag{14}$$

$$Precision = TP / TP + FP \tag{15}$$

$$Recall = TP / TP + FN \tag{16}$$

The table below displays the accuracy, recall, F1 score, and precision analysis of each model trained on a dataset for categorising lung sounds. The recall, F1 score, and Precision values are displayed for various train-test data splits, as percentages for each model.

Table 1. Classification Results of deep learning architectures, like Precision, Recall, and F1-Score of each model

Model	pulmonary	precision	Recall	F1-score
CNN	COPD	0.58	0.78	0.67
	Bronchiolitis	0.11	0.40	0.17
	Pneumonia	0.94	0.99	0.96
	URTI	0.80	0.29	0.42
	Healthy	0.11	0.40	0.17
LSTM	COPD	1.00	1.00	1.00
	Bronchiolitis	0.75	1.00	0.86
	Pneumonia	0.75	0.75	0.75
	URTI	1.00	0.75	0.86
	Healthy	0.82	0.69	0.75
CRNN	COPD	1.00	0.90	0.95
	Bronchiolitis	0.86	1.00	0.92
	Pneumonia	0.94	0.99	0.86
	URTI	1.00	0.75	0.86
	Healthy	0.82	0.69	0.75
VGG16	URTI	0.00	0.00	0.00
	Healthy	0.00	0.00	0.00
	Asthma	0.00	0.00	0.00
	COPD	0.83	1.00	0.91
	LRTI	0.00	0.00	0.00
	Bronchiolitis	0.06	0.04	0.04
	Pneumonia	0.02	0.04	0.03
ResNet50	URTI	0.00	0.00	0.00
	Healthy	0.00	0.00	0.00
	Asthma	0.00	0.00	0.00
	COPD	0.83	0.87	0.85
	LRTI	0.00	0.00	0.00
	Bronchiolitis	0.06	0.04	0.04
	Pneumonia	0.02	0.04	0.03

The outcomes indicate that the performance of 5 models, CNN, VGG16, ResNet50, CRNN, and LSTML, fluctuates depending on the task at hand. CNN showcased its effectiveness in data recognition by achieving the highest accuracy rate of 98% shown in Figure 14. Contrastingly, VGG16 achieved an accuracy of only 83%, shown in Figure 11, significantly lower than that of CNN. Despite its reputation for simplicity and effectiveness, the VGG16 model performed notably poorer in this specific task. While CNN excelled with an accuracy of 98%, VGG16 and ResNet50 also achieved respectable results of 83% and 95%, respectively. These findings highlight the importance of model selection and optimization in maximizing results in machine learning projects. The LSTML model obtained an accuracy of 85%, as shown in Figure 8, while the CRNN model achieved an accuracy of 87%. When all the models are compared, the CNN model outperforms all the others, obtaining 98% accuracy.

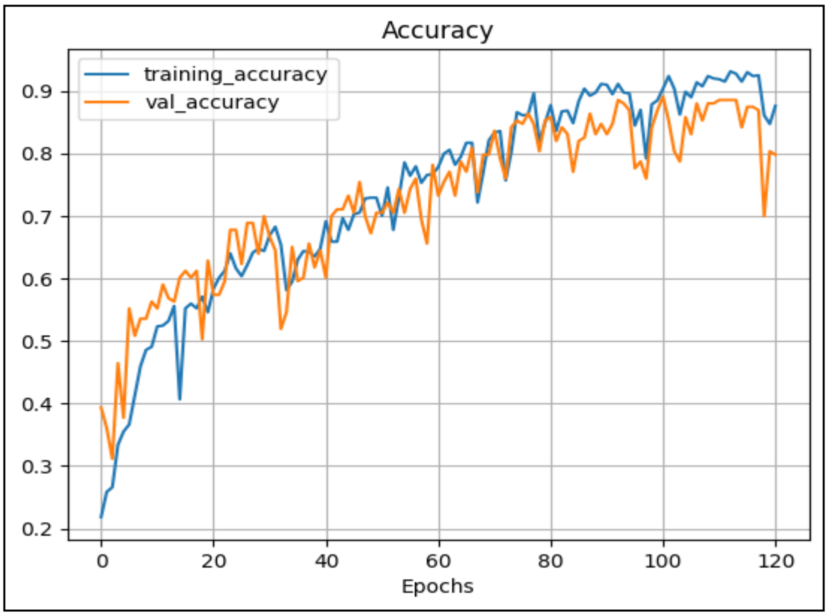

Figure 8. Training and validation accuracy of the LSTM model

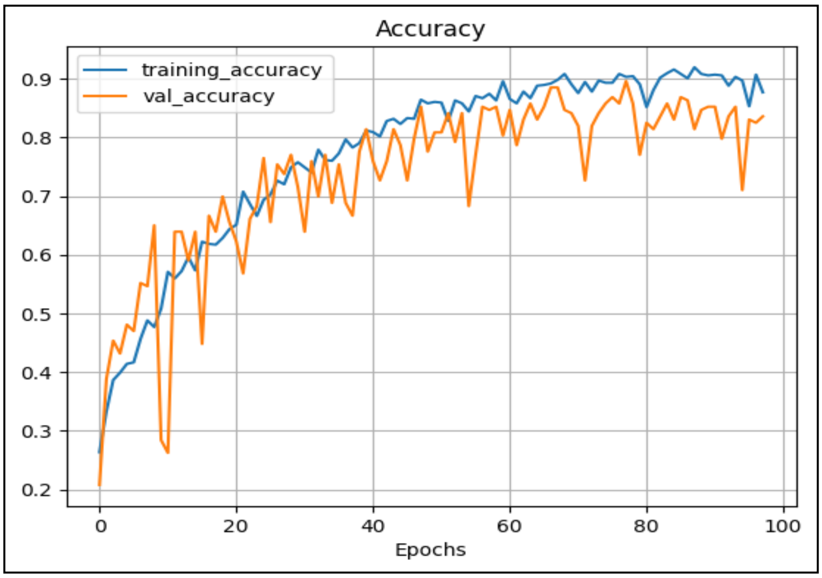

Figure 9. Train and Validation accuracy of CRNN model

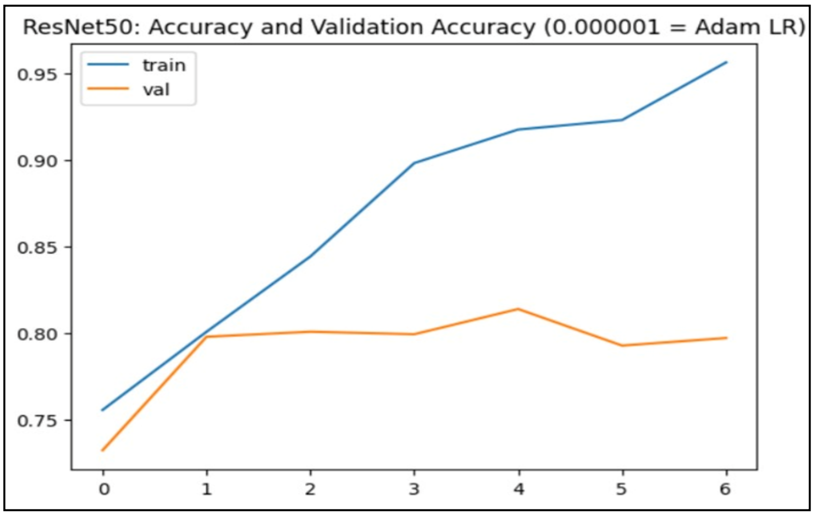

Figure 10. Training and validation accuracy of the ResNet50 model

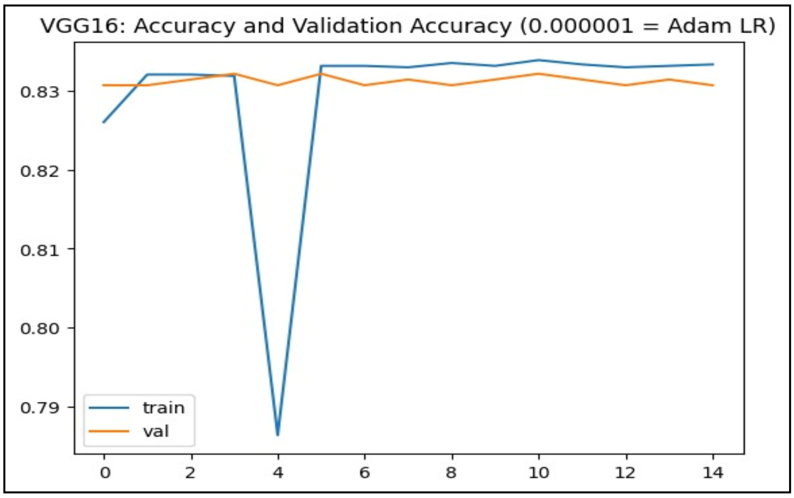

Figure 11. Training and validation accuracy of the VGG16 model

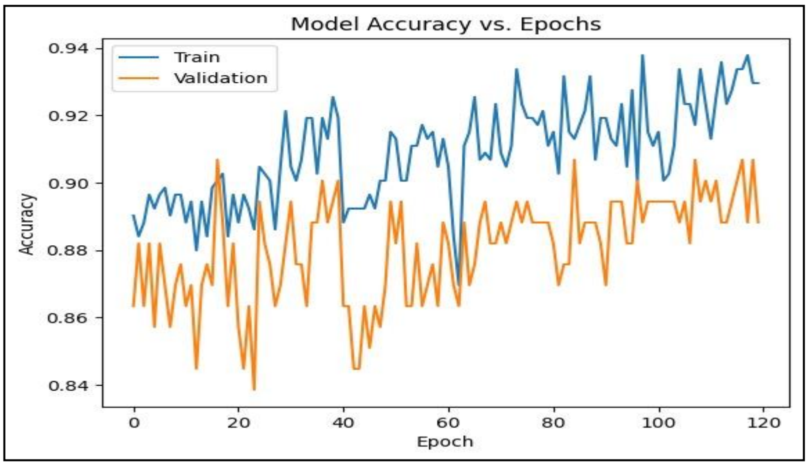

Figure 12. Training and validation accuracy of the CNN model

Conclusion

Respiratory illnesses have long had a substantial impact on human health. Proper diagnosis and early treatment can significantly enhance patients' quality of life. In this study, focusing on lung sound classification, convolutional neural networks (CNNs) have shown promising outcomes in accurately categorizing respiratory sounds. The present paper proposes a method to recognize the sound of the respiratory system. The approach is especially beneficial for categorizing respiratory disorders. The identification approach first begins processing the respiratory sounds for respiratory cycle features, then converts the cycle to a spectrogram, and then classifies it with a traditional deep learning network. Finally, an analysis is done on the given results. By leveraging transfer learning and diversifying datasets, CNN, CRNN, and LSTM models can become more versatile, ultimately boosting their accuracy in identifying respiratory issues. This project sets the groundwork for cutting-edge diagnostic tools in respiratory healthcare and represents a potential advancement in early detection and monitoring.

Acknowledgments

The authors express their sincere gratitude to Christ University, Bengaluru, Presidency College, Bengaluru, and Manipal Academy of Higher Education, Manipal, for their valuable support in facilitating this research. Their academic environment and infrastructure greatly contributed to the successful completion of the study.

Conflict of interest

The authors declare no potential conflict of interest regarding the publication of this work.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Ali, Shams Nafisa, et al. "An End-to-end Deep Learning Framework for Real-Time Denoising of Heart Sounds for Effective Computer-Aided Auscultation in Unseen Noise." *Authorea Preprints* (2023).

Anupama, H. S., K. R. Pradeep, G. Shreeya, Pratiksha Rao, and S. K. Tejasvi. "Detection of Chronic Lung Disorders using Deep Learning." In *2022 Fourth International Conference on Cognitive Computing and Information Processing (CCIP)*, pp. 1-5. IEEE, 2022.

Bapa, Aditya, Omkar Bandgar, Arnav Ekapure, and Jignesh Sisodia. "Respiratory disorder classification based on lung auscultation using MFCC, Mel Spectrogram and Chroma STFT." In *2023 International Conference on Artificial Intelligence and Applications (ICAIA) Alliance Technology Conference (ATCON-1)*, pp. 1-7. IEEE, 2023.

Khanaghavalle GR, Manoj A, Karthikeyan J, Murali R. A Deep Learning Framework for Multiclass Categorization of Pulmonary Diseases. In *2023 World Conference on Communication & Computing (WCONF)*, pp. 1-6. IEEE, 2023.

Kilic, M., Barua, P. D., Keles, T., Yildiz, A. M., Tuncer, I., Dogan, S., ... & Acharya, U. R. (2024). GCLP: An automated asthma detection model based on global chaotic logistic pattern using cough sounds. *Engineering Applications of Artificial Intelligence*, 127, 107184.

Koshta, Vaibhav, Bikesh Kumar Singh, Ajoy K. Behera, and Ranganath T. Ganga. "Classification of Asthma, COPD and Healthy Lung Sounds Using Fourier Bessel Series Expansion in Machine Learning and Deep Learning Paradigm." In *2023 11th International Conference on Intelligent Systems and Embedded Design (ISED)*, pp. 1-6. IEEE, 2023.

Kumar, Dhirendra. "Multi Spectral Feature Extraction to Improve Lung Sound Classification using CNN." In *2023 10th International Conference on Signal Processing and Integrated Networks (SPIN)*, pp. 186-191. IEEE, 2023.

Ma, W. B., Deng, X. Y., Yang, Y., & Fang, W. C. (2022). An effective lung sound classification system for respiratory disease diagnosis using DenseNet CNN model with sound preprocessing engine. In *2022 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pp. 218-222. IEEE.

Pham, Lam, Huy Phan, Ramaswamy Palaniappan, Alfred Mertins, and Ian McLoughlin. "CNN-MoE based framework for classification of respiratory anomalies and lung disease detection." *IEEE Journal of Biomedical and Health Informatics*, 25(8), 2938–2947, 2021.

Phettom, R., Theera-Umpon, N., & Auephanwiriyakul, S. (2023). Automatic Identification of Abnormal Lung Sounds Using Time-Frequency Analysis and CNN. In *2023 ICITEE*, pp. 1-6. IEEE.

Roy, Arka, and Udit Satija. "RDLINet: A Novel Lightweight Inception Network for Respiratory Disease Classification Using Lung Sounds." *IEEE Transactions on Instrumentation and Measurement* (2023).

Shuvo, S. B., Ali, S. N., Swapnil, S. I., Hasan, T., & Bhuiyan, M. I. H. (2020). 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*, 25(7), 2595–2603.

Tariq, Zeenat, Sayed Khushal Shah, and Yugyung Lee. "Lung disease classification using deep convolutional neural network." In *2019 IEEE BIBM*, pp. 732-735. IEEE, 2019.

Tariq, Zeenat, Sayed Khushal Shah, and Yugyung Lee. "Multimodal lung disease classification using deep convolutional neural network." In *2020 IEEE BIBM*, pp. 2530-2537. IEEE, 2020.

Tsai, K. H., Wang, W. C., Cheng, C. H., Tsai, C. Y., Wang, J. K., Lin, T. H., ... & Tsao, Y. (2020). Blind monaural source separation on heart and lung sounds based on periodic-coded deep autoencoder. *IEEE Journal of Biomedical and Health Informatics*, 24(11), 3203–3214.

Wang, F., Yuan, X., & Meng, B. (2023). Classification of Abnormal Lung Sounds Using Deep Learning. In *2023 ICSIP*, pp. 506-510. IEEE.

Yamashita, M. (2021). Classification Between Normal and Abnormal Respiration Using Ergodic HMM for Intermittent Abnormal Sounds. In *2021 European Signal Processing Conference (EUSIPCO)*, pp. 1187-1190. IEEE.

Zakaria, Neili, Fezari Mohamed, Redjati Abdelghani, and Kenneth Sundaraj. "VGG16, ResNet-50, and GoogLeNet deep learning architecture for breathing sound classification: a comparative study." In *2021 AI-CSP Conference*, pp. 1-6. IEEE, 2021.

Bibliographic information of this paper for citing:

K S, Hemanth; Naik T, Harisha; Kartik, N; kumar, N Nanda; Senthilkumar, S & R, Ramya (2026). Attentional Deep Learning with Inverse Transform Sampling for Robust Respiratory Sound Classification. *Journal of Information Technology Management*, 18 (1), 123-140.
https://doi.org/10.22059/jitm.2026.1062557