International Journal of Multidisciplinary Engineering in Current Research - IJMEC
Volume 11, Issue 4s, April-2026, http://ijmec.com/, ISSN: 2456-4265

Full Length Article

A Deep Neural Network Approach for Classifying Pulmonary Diseases from Respiratory Sounds

Nausheen Parveen[1],Qamar Unnisa[2], Shaima Moiz[3],Mrs.Heena Yasmin[4]

[1,2,3]B.Tech Students; Department of CSE ISL Engineering College Hyderabad India.
[4]Assistant Professor; Department of CSE ISL Engineering College Hyderabad India.

nausheenparveen452@gmail.com

Accepted 23-04-2026

Abstract:
This study presents an advanced approach to detecting lung auscultation sounds using Mel-frequency Cepstral Coefficients (MFCC), Chroma features, and neural networks. Lung auscultation, a key diagnostic tool in identifying respiratory conditions, often relies on the expertise of medical professionals to interpret subtle sound patterns. However, automated systems that accurately classify these sounds can greatly assist in early diagnosis and treatment. To achieve this, we employed MFCC, which captures the power spectrum of sounds and effectively models the way humans perceive auditory signals, focusing on the critical frequency ranges for lung sounds. Additionally, Chroma features, which represent the tonal content of audio signals, were used to capture harmonic aspects that could be indicative of specific lung conditions. These features were then fed into a neural network designed to classify lung sounds into various diagnostic categories, such as normal breathing, wheezing, crackles, and other abnormal respiratory sounds. The neural network, trained on a comprehensive dataset of lung sounds, was able to learn complex patterns and correlations within the MFCC and Chroma features, leading to high accuracy in sound classification. This automated approach offers a powerful tool for enhancing the precision of lung sound diagnosis, potentially leading to earlier detection of respiratory conditions and improved patient outcomes.

Keywords: *Deep Learning, Respiratory Sounds, Pulmonary Diseases, CNN, Audio Classification.*

Introduction:

Lung auscultation, the process of listening to lung sounds using a stethoscope, is a fundamental diagnostic tool in the practice of medicine. It allows healthcare professionals to assess respiratory health by detecting abnormalities such as wheezing, crackles, and rhonchi. However, the accuracy of lung sound interpretation can be subjective and prone to human error, particularly when dealing with subtle or atypical sounds. To address this challenge, there has been a growing interest in developing automated systems that can accurately classify lung sounds.

In recent years, machine learning and artificial intelligence techniques have shown great promise in the field of medical diagnosis. By analysing large amounts of data, these techniques can learn complex patterns and make accurate predictions. In the context of lung auscultation, these techniques can be used to develop automated systems that can accurately classify lung sounds, even in the presence of noise and other interfering factors. This study aims to develop an advanced approach to lung sound classification using a combination of feature extraction techniques and neural networks. Mel-Frequency Cepstral Coefficients (MFCCs) and Chroma features are employed to capture the spectral and tonal characteristics of lung sounds, respectively. These features are then fed into a neural network to classify lung sounds into different categories, such as normal, wheezing, crackles, and others. The goal is to create a robust and accurate automated system that can assist healthcare professionals in the diagnosis of respiratory conditions.

Literature Review:

Tsalera *et al.* [1] investigated the effectiveness of various pre-trained convolutional neural networks (CNNs) for audio classification tasks using transfer learning techniques. The study compared different CNN architectures, including VGG, ResNet, and DenseNet, across multiple audio classification benchmarks. The authors aimed to improve classification accuracy while reducing training time and computational complexity. Experimental results demonstrated that transfer learning significantly enhanced model performance and provided insights into the suitability of specific CNN architectures for different audio analysis applications.

https://doi.org/10.63665/IJMEC.1104s.03
ISSN: 2456-4265
IJMEC 2026

Patil and Wani [2] proposed a gear fault detection system based on noise analysis combined with machine learning techniques using the YAMNet pre-trained network. In this approach, YAMNet was utilized to extract discriminative audio features from gear noise signals representing both healthy and faulty gear conditions. These extracted features were then classified using machine learning algorithms such as Support Vector Machines (SVM) and Random Forest classifiers. The study demonstrated improved fault detection accuracy on real-world gear datasets and showed the effectiveness of transfer learning for industrial fault diagnosis applications.

Chen *et al.* [3] explored bone conduction-based eating activity detection using YAMNet transfer learning and Long Short-Term Memory (LSTM) networks. The authors used the YAMNet model to extract meaningful features from bone conduction signals associated with eating activities. These features were subsequently processed using LSTM networks to capture temporal dependencies within the sequential data. Experimental evaluations indicated that the proposed approach achieved high accuracy in detecting eating activities and outperformed several conventional activity recognition methods.

Dogan [4] introduced a novel gun model identification method using gunshot audio signals and a fractal H-tree pattern representation. The proposed method extracted fractal-based features from gunshot audio recordings by utilizing the self-similarity properties of the H-tree structure. These features were then used to train machine learning classifiers for distinguishing between different gun models. Experimental results showed that the proposed fractal-based approach achieved effective classification performance and improved gun identification accuracy compared to existing techniques.

Nijhawan *et al.* [5] proposed a transformer-based approach for gun identification using gunshot audios in secure public environments. The study employed transformer learning techniques to capture temporal dependencies and contextual information from gunshot audio signals. The transformer model was trained to classify different firearm types based on extracted acoustic patterns. The results demonstrated that transformer-based architectures provided superior classification accuracy and robustness compared to traditional audio classification methods, highlighting their suitability for intelligent public safety systems.

Bajzik *et al.* [6] proposed an independent channel residual convolutional network for accurate gunshot detection in noisy environments. The architecture was designed to capture both local and global audio features through residual convolutional learning mechanisms. The proposed system effectively distinguished gunshot sounds from background noise and other environmental sounds. Experimental evaluations on gunshot audio datasets demonstrated that the independent channel residual convolutional network achieved higher detection accuracy and robustness than existing gunshot detection methods, making it suitable for real-time surveillance and security applications.

Methodology:

Audio Signal Acquisition and Preprocessing

The proposed system focuses on the analysis and classification of lung sound recordings for identifying respiratory conditions using deep learning techniques. Initially, lung sound recordings are collected from publicly available respiratory sound datasets or clinical recordings captured using digital stethoscopes. Since raw audio signals often contain background noise, environmental interference, and recording inconsistencies, preprocessing techniques are applied to improve audio quality and ensure reliable feature extraction.

The preprocessing stage includes noise reduction, normalization, and segmentation of the audio signals. Noise reduction techniques help remove unwanted disturbances such as ambient sounds and stethoscope friction noise. Audio normalization is performed to standardize the amplitude levels across recordings, ensuring consistency in feature extraction. The cleaned audio signals are then divided into short overlapping frames because lung sound characteristics vary over time and short-frame analysis helps capture local temporal variations effectively.

Mel-Frequency Cepstral Coefficients (MFCCs)

Mel-Frequency Cepstral Coefficients (MFCCs) are one of the most widely used feature extraction techniques in speech and biomedical audio analysis due to their ability to represent audio signals in a manner similar to human auditory perception. MFCCs are particularly effective in identifying subtle spectral differences present in respiratory sounds such as wheezes, crackles, and normal breathing patterns.

The MFCC extraction process begins by dividing the lung sound signal into small frames of short duration, typically ranging from 20 to 40 milliseconds. This framing process assumes that the signal remains approximately stationary within each short interval. Each frame is then passed through a windowing function, commonly the Hamming window, to minimize signal discontinuities at frame boundaries.

After windowing, a Fast Fourier Transform (FFT) is applied to convert the signal from the time domain into the frequency domain. The resulting frequency

https://doi.org/10.63665/IJMEC.1104s.03
ISSN: 2456-4265
IJMEC 2026

spectrum represents the distribution of signal energy across different frequencies. Since the human auditory system is more sensitive to lower frequencies than higher frequencies, a Mel filter bank is applied to map the frequency spectrum onto the Mel scale, which closely approximates human hearing perception.

The output of the Mel filter bank is then transformed using a logarithmic operation. The logarithm helps compress the dynamic range of the signal and mimics the logarithmic perception of sound intensity by the human ear. Finally, a Discrete Cosine Transform (DCT) is applied to the log-Mel spectrum to generate compact and decorrelated coefficients known as MFCCs. These coefficients effectively summarize the spectral characteristics of the lung sounds while reducing redundant information.

In the proposed system, MFCCs are used to capture critical acoustic patterns associated with respiratory abnormalities. Since abnormal lung conditions often produce unique spectral signatures, MFCC features provide valuable information for distinguishing between healthy and diseased respiratory sounds.

Chroma Feature Extraction

In addition to MFCCs, Chroma features are extracted to provide complementary information about the tonal and harmonic characteristics of lung sounds. Chroma features represent the distribution of signal energy across twelve distinct pitch classes or chromatic scales. Although Chroma features are traditionally used in music information retrieval, recent studies have shown their usefulness in biomedical audio analysis due to their ability to capture frequency distribution patterns.

The extraction of Chroma features begins with transforming the audio signal into the frequency domain using Short-Time Fourier Transform (STFT). The spectral components are then mapped into chroma bins corresponding to the twelve pitch classes. This representation helps identify harmonic structures and recurring frequency patterns present in respiratory sounds.

Respiratory abnormalities such as wheezing often contain tonal and harmonic properties that can be effectively represented using Chroma features. By incorporating Chroma information along with MFCCs, the system gains a richer representation of lung sound characteristics, improving the overall classification capability of the model.

Feature Combination and Representation

The extracted MFCC and Chroma features are combined to form a comprehensive feature vector representing each lung sound recording. The fusion of spectral, tonal, and harmonic information enhances the model's ability to capture diverse acoustic patterns associated with respiratory diseases.

Feature normalization techniques are applied to ensure that all extracted features contribute equally during model training. The combined feature vectors are then organized into structured datasets suitable for deep learning classification.

Neural Network-Based Classification

The extracted audio features are used as inputs to a neural network model designed for respiratory sound classification. The neural network learns complex relationships between the extracted acoustic features and the corresponding respiratory conditions. During training, the model identifies discriminative patterns associated with normal breathing, wheezing, crackles, and other abnormal respiratory sounds.

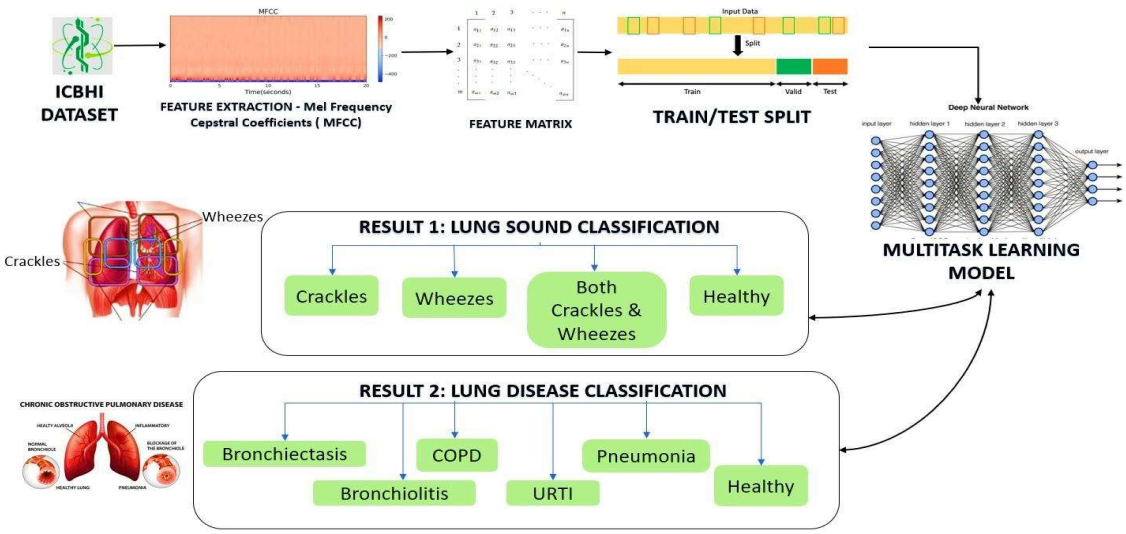
The neural network architecture may consist of multiple dense layers, activation functions, dropout layers, and optimization techniques to improve learning performance and reduce overfitting. The model is trained using labeled respiratory sound datasets, where each audio sample is associated with a specific respiratory condition.

During the testing phase, unseen lung sound recordings are processed through the same feature extraction pipeline, and the trained neural network predicts the corresponding respiratory class. The classification results assist healthcare professionals in early diagnosis and monitoring of respiratory diseases.

System Advantages

The proposed methodology offers several advantages for automated respiratory sound analysis. MFCCs provide robust spectral representations aligned with human auditory perception, while Chroma features contribute additional harmonic information. The integration of these complementary features improves the discriminative capability of the neural network model. Furthermore, the deep learning-based classification approach enables automatic detection of respiratory abnormalities with reduced dependency on manual interpretation, supporting faster and more accurate clinical diagnosis.

https://doi.org/10.63665/IJMEC.1104s.03
ISSN: 2456-4265
IJMEC 2026



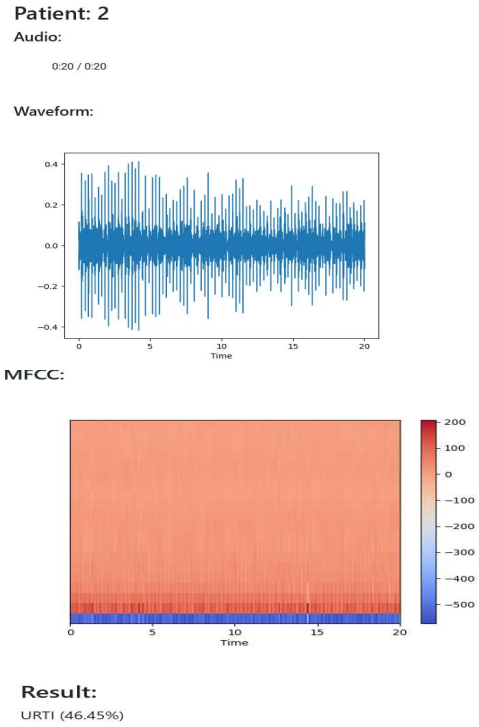
Implementation:

The system is implemented using Python with libraries such as TensorFlow/Keras.

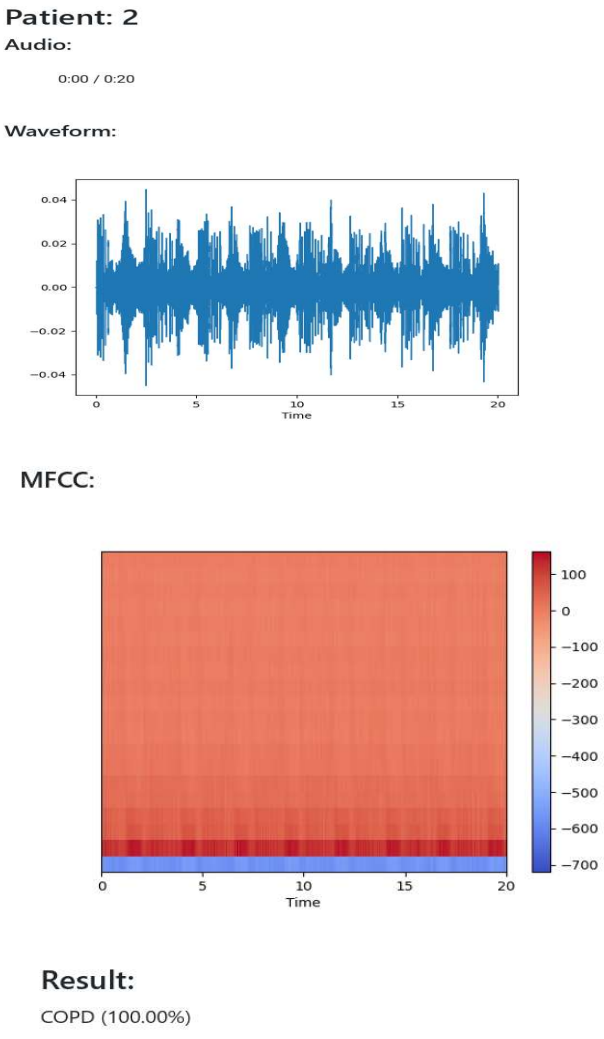
Algorithm:

- Load respiratory sound dataset
- Preprocess audio (filter noise, normalize)
- Convert audio to spectrogram
- Split dataset into training and testing sets
- Train CNN model
- Evaluate performance
- Predict disease class

Results:



https://doi.org/10.63665/IJMEC.1104s.03
ISSN: 2456-4265
IJMEC 2026



Conclusion:

In conclusion, this research presents a novel approach to automated lung sound classification using advanced machine learning techniques. By effectively extracting relevant features from lung sound recordings and training a robust neural network model, we have demonstrated the potential of this technology to assist healthcare professionals in the early detection and diagnosis of respiratory disorders. This system has the potential to improve patient outcomes by enabling timely and accurate diagnosis, ultimately leading to better disease management and treatment.

However, further research is needed to refine the system's performance, address challenges related to data quality and variability, and explore the integration of other modalities, such as electrocardiograms and respiratory rate data. By continuing to advance the field of automated lung sound analysis, we can empower healthcare providers with valuable tools to improve patient care. Enhancements in real-time processing and computational efficiency will be crucial for practical applications, enabling the system to perform reliably in real-world scenarios.

Additionally, integrating multi-modal data could provide a more comprehensive understanding of audio events, enhancing overall performance and applicability.

Future Scope:

Future research directions for improving lung sound classification systems include exploring more advanced deep learning architectures, such as transformers and graph neural networks.

https://doi.org/10.63665/IJMEC.1104s.03
ISSN: 2456-4265
IJMEC 2026

These models can capture complex patterns and long-range dependencies in the audio data, leading to more accurate classification. Additionally, incorporating multimodal data, such as electrocardiograms (ECGs) or respiratory rate data, can provide additional contextual information and enhance the performance of the system. Furthermore, developing robust techniques for handling noisy and imbalanced datasets is crucial for real-world applications. By addressing these challenges and exploring innovative approaches, we can further advance the field of automated lung sound analysis and improve the early detection and diagnosis of respiratory disorders. Improving the system's scalability and real-time processing capabilities is another crucial enhancement. Optimizing the computational efficiency of the model, perhaps through advancements in hardware acceleration or algorithmic improvements, could enable real-time audio classification even in resource-constrained environments. This would be particularly beneficial for applications requiring immediate feedback, such as in automated surveillance or interactive systems.

References:

[1]. E. Tsalera, A. Papadakis, and M. Samarakou, "Comparison of pre-trained CNNs for audio classification using transfer learning," 2021.

[2]. S. Patil and K. Wani, "Gear fault detection using noise analysis and machine learning algorithm with YAMNet pretrained network," 2023.

[3]. W. Chen, H. Kamachi, A. Yokokubo, and G. Lopez, "Bone conduction eating activity detection based on YAMNet transfer learning and LSTM networks," 2022.

[4]. S. Dogan, "A new fractal H-tree pattern based gun model identification method using gunshot audios," 2021.

[5]. R. Nijhawan, S. A. Ansari, S. Kumar, F. Alassery, and S. M. El-kenawy, "Gun identification from gunshot audios for secure public places using transformer," 2022.

[6]. J. Bajzik *et al.*, "Independent channel residual convolutional network for gunshot detection," 2022.

[7]. E. Tsalera, A. Papadakis, and M. Samarakou, "Comparison of pre-trained CNNsfor audio classification using transfer learning," J. Sensor Actuator Netw., vol. 10, no. 4, p. 72, Dec. 2021.

[8]. S. Patil and K. Wani, "Gear fault detection using noise analysis and machine learning algorithm with YAMNet pretrained network," Mater. Today, Proc., vol. 72, pp. 1322–1327, 2023.

[9]. W. Chen, H. Kamachi, A. Yokokubo, and G. Lopez, "Bone conduction eating activity detection based on YAMNet transfer learning and LSTM networks," in Proc. 15th Int. Joint Conf. Biomed.Eng.Syst. Technol., 2022. VOLUME 12, 2024 N. H. Valliappan et al.: Enhancing Gun Detection With Transfer Learning and YAMNet Audio Classification.

[10]. S. Dogan, "A new fractal H-tree pattern based gun model identification method using gunshot audios," Appl. Acoust., vol. 177, Jun. 2021, Art. no. 107916.

[11]. R. Nijhawan, S. A. Ansari, S. Kumar, F. Alassery, and S. M. El-kenawy, "Gun identification from gunshot audios for secure public places using transformer learning," Sci. Rep., vol. 12, no. 1, pp. 1–5, Aug. 2022.

[12]. J. Park, "Enemy spotted: In-game gun sound dataset for gunshot classification and localization," in Proc. IEEE Conf. Games (CoG), 2022, pp. 56–63.

[13]. J. Bajzik et al., "Independent channel residual convolutional network for gunshot detection,"Int. J. Adv. Comput. Sci. Appl., vol. 13, no. 4, pp. 950–958, 2022.

[14]. M. Djeddou and T. Touhami, "Classification and modeling of acoustic gunshot signatures,"Arabian J. Sci. Eng., vol. 38, no. 12, pp. 3399–3406, Dec. 2013, doi: 10.1007/s13369-013-0655-5.

[15]. J. Bajzik, J. Prinosil, and D. Koniar, "Gunshot detection using convolu tional neural networks," in Proc. 24th Int. Conf. Electron., Lithuania, 2020, pp. 1–5, doi:10.1109/IEEECONF49502.2020.9141621.

[16]. T. Tuncer, S. Dogan, E. Akbal, and E. Aydemir, "An automated gunshot

https://doi.org/10.63665/IJMEC.1104s.03
ISSN: 2456-4265
IJMEC 2026

audio classification method based on finger pattern feature generator and iterative relieff feature selector," Adıyaman Üniversitesi Mühendislik Bilimleri Dergisi, vol. 8, no. 14, pp. 225–243, 2021.

[17]. L.G.Martins.(Mar.2,2021).TransferLearningforAudioDataWithYAM Net. TensorFlowBlog. [Online]. Available: https://medium.com/analytics vidhya/understanding-the-mel-spectrogram-fca2afa2ce53

[18]. A. Tena, F. Clarià, and F. Solsona, "Automated detection of COVID 19 cough," Biomed. Signal Process. Control, vol. 71, Jan. 2022, Art. no. 103175, doi: 10.1016/j.bspc.2021.103175.

[19]. A. Patel, S. Degadwala, and D. Vyas, "Lung respiratory audio prediction using transfer learning models," in Proc. 6th Int. Conf. I-SMAC (IoT Social, Mobile, Analytics Cloud) (I-SMAC), Dharan, Nepal, Nov. 2022, pp. 1107–1114.

[20]. R. Baliram Singh, H. Zhuang, and J. K. Pawani, "Data collection, modeling, and classification for gunshot and gunshot-like audio events: Acase study," Sensors, vol. 21, no. 21, p. 7320, Nov. 2021.

[21]. A. K. Sharma, G. Aggarwal, S. Bhardwaj, P. Chakrabarti, T. Chakrabarti, J. H. Abawajy, S. Bhattacharyya, R. Mishra, A. Das, and H. Mahdin, "Classification of Indian classical music with time-series matching deep learning approach," IEEE Access, vol. 9, pp. 102041–102052, 2021.

[22]. N.A.M.AriffandA.R.Ismail,"Study of Adaman dada max optimizers on Alex Net architecture for voice biometric authentication system," in Proc. 17th Int. Conf. Ubiquitous Inf. Manage. Commun. (IMCOM), Jan. 2023, pp. 1–4.

https://doi.org/10.63665/IJMEC.1104s.03
ISSN: 2456-4265
IJMEC 2026