Classification and Recognition of Lung Sounds Based on BI-ResNet Model

[1]D.Shyam Sundar, [2]G.Kulaeep Nayyar, [3]E.Barath Kumar, [4]Ms.Pooja Kulkarni
[1,2,3] UG Scholars, [4] Assistant Prof
[1,2,3,4] Department of Computer Science Engineering,
[1,2,3,4] Guru Nanak Institutions Technical Campus, Hyderabad, Telangana, India.

ABSTRACT:

This research introduces a novel method for identifying lung sounds using a combination of Mel-Frequency Cepstral Coefficients (MFCC), Chroma features, and neural networks. Lung auscultation plays a crucial role in diagnosing respiratory illnesses, yet it typically depends on a clinician's skill to detect and interpret subtle audio cues. To support and enhance this diagnostic process, we developed an automated system capable of accurately recognizing and categorizing lung sounds. MFCCs were utilized to analyze the spectral characteristics of audio signals, mimicking how the human ear processes sound, with emphasis on frequency bands relevant to respiratory acoustics. In parallel, Chroma features were extracted to capture tonal and harmonic elements that may reflect specific respiratory anomalies. These combined audio features were input into a neural network trained to distinguish between different types of lung sounds, such as normal respiration, wheezing, crackles, and other abnormal patterns. By learning intricate relationships within the MFCC and Chroma data, the model achieved high classification accuracy. This approach offers a promising solution for improving the reliability and timeliness of respiratory disease diagnosis, ultimately contributing to better patient care and outcomes.

1 INTRODUCTION

Lung auscultation—the act of listening to respiratory sounds through a stethoscope—is a core component of medical diagnostics. It enables clinicians to evaluate a patient's respiratory condition by identifying irregularities such as wheezes, crackles, and rhonchi. Despite its importance, the interpretation of these sounds can be highly subjective and may vary between practitioners, especially when dealing with faint or unusual sound patterns. To overcome this limitation, there is increasing interest in creating automated systems capable of reliably analyzing and classifying lung sounds.

Advancements in machine learning and artificial intelligence have opened new avenues in medical diagnostics. These technologies are capable of processing large datasets to detect intricate patterns and make precise predictions, offering a promising solution for improving the accuracy and consistency of respiratory sound analysis.

1.1 OBJECTIVE:

The main goal of this study is to design a reliable and precise automated system for classifying lung sounds, with the intent of supporting healthcare professionals in the early identification and diagnosis of respiratory conditions. This system leverages advanced machine learning models—particularly neural networks—along with effective feature extraction techniques, including Mel-Frequency Cepstral Coefficients (MFCC) and Chroma features. By doing so, the proposed model aims to accurately distinguish between different types of lung sounds, such as normal breathing, wheezes, crackles, and other abnormal respiratory patterns. This tool is expected to assist clinicians in making more accurate diagnoses, ultimately contributing to better patient care and health outcomes.

1.2 SCOPE OF THE WORK:

This research focuses on designing and assessing an innovative automated system for lung sound classification. The central aim is to build a reliable and precise tool that can support medical professionals in the early identification and diagnosis of respiratory conditions. By advancing the field of automated lung sound analysis, this work seeks to enhance the accuracy and efficiency of diagnosing respiratory diseases.

1.3 PROBLEM STATEMENT:

Accurate classification and recognition of lung sounds are critical for the early diagnosis and management of respiratory diseases. Traditional auscultation methods rely heavily on the subjective interpretation of healthcare professionals, which can lead to inconsistencies and diagnostic errors, particularly in cases involving subtle or complex respiratory sounds. While machine learning techniques have shown potential in automating this process, many existing models struggle with effectively capturing both local and global features of lung sound signals.

To address this issue, this research proposes the use of a Bidirectional Residual Network (Bi-ResNet) model for lung sound classification and recognition. The Bi-ResNet architecture is designed to leverage the strengths of residual learning and bidirectional information flow, allowing for deeper feature extraction and better temporal representation of lung sound patterns. The objective is to develop a robust and accurate system capable of distinguishing between various lung sound categories—such as normal, wheeze, crackle, and other abnormal patterns—thereby improving diagnostic reliability and supporting clinical decision-making.

Technical Challenge Focused

The development of an effective lung sound classification system based on the Bi-ResNet model presents several technical challenges:

1. **Noisy and Non-Stationary Signals:** Lung sound recordings often contain background noise, artifacts from patient movement, and overlapping heart sounds, making it difficult to isolate and classify respiratory events accurately.
2. **Imbalanced and Limited Datasets:** Publicly available lung sound datasets often suffer from class imbalance (e.g., fewer samples of rare conditions like stridor) and limited annotated recordings, which can hinder the generalization performance of deep learning models like Bi-ResNet.
3. **Feature Representation of Time-Frequency Data:** Converting raw audio into meaningful representations (e.g., MFCCs, spectrograms) that effectively capture both spectral and temporal characteristics is crucial. Ensuring compatibility and optimal performance of these features with the Bi-ResNet architecture requires careful preprocessing and design.
4. **Model Complexity and Overfitting:** Bi-ResNet architectures are deeper and more complex than standard CNNs. While this allows for richer feature extraction, it also increases the risk of overfitting, especially when training on small or noisy datasets.
5. **Bidirectional Contextual Learning:** Effectively leveraging bidirectional pathways within the ResNet structure for time-dependent lung sound data is a non-trivial challenge. It requires architectural adaptations and careful tuning to ensure that both past and future contextual information is utilized without introducing vanishing gradient issues.
6. **Real-Time Implementation and Resource Constraints:** Deploying the Bi-ResNet model in real-world settings, such as mobile health devices or low-power diagnostic tools, demands optimization for speed, memory usage, and inference efficiency without sacrificing classification accuracy.

1.4 EXISTING SYSTEM:

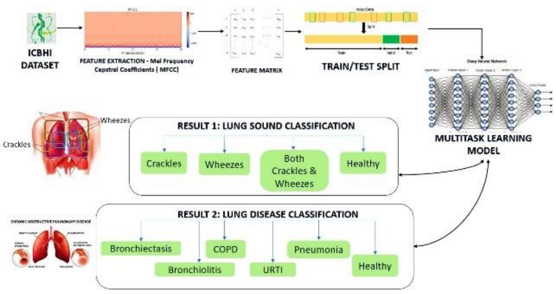
- The **Bidirectional Residual Network (Bi-ResNet)** is an enhanced architecture based on the traditional Residual Network (ResNet), designed to integrate bidirectional data processing within the residual learning framework. This approach retains the core advantages of ResNet while adding features that make it especially effective for tasks involving sequential or time-dependent data.
- Similar to ResNet, the Bi-ResNet model employs **skip (residual) connections** that facilitate better gradient flow during training. These connections help avoid issues like vanishing gradients and allow for the successful training of very deep networks by learning residual mappings instead of direct transformations.
- What sets Bi-ResNet apart is its **bidirectional processing capability**, which means it analyzes input sequences in both forward and reverse directions. This is particularly beneficial for applications where understanding both preceding and succeeding context enhances performance, such as time-series analysis or audio signal processing. By capturing dependencies from both directions, Bi-ResNet is better equipped to model complex temporal relationships and subtle sequential patterns.

.

1.4.1 DISADVANTAGES:

- Computationally intensive
- Prone to overfitting, especially with limited data.
- Difficult to interpret and explain the internal decision-making process.

1.5SYSTEM ARCHITECTURE:



1.6 PROPOSED SYSTEM

- Mel-Frequency Cepstral Coefficients (MFCCs) are widely used features in the field of audio and speech processing due to their effectiveness in capturing perceptually relevant information. They are based on the Mel scale, which reflects how the human auditory system perceives frequency, placing more emphasis on lower frequencies.
- The extraction of MFCCs starts by segmenting the audio signal into short time frames. Each of these frames is then transformed into the frequency domain using the Fourier Transform. To simulate human hearing, a Mel-scale filter bank is applied, which boosts sensitivity to lower-frequency components while reducing the influence of higher frequencies. This process results in a compact set of features that effectively represent the spectral characteristics of the original audio signal.
  - To accomplish this, we utilized MFCCs, which analyze the power spectrum of audio signals and closely replicate human auditory perception by emphasizing the frequency ranges most relevant to lung sounds.

1.6.1 ADVANTAGES

- Chroma features effectively capture the harmonic and tonal content of audio signals
- Neural networks are adaptable and can be tailored to different tasks and datasets.
- MFCCs are versatile and can be applied to various audio processing tasks.

2 PROJECT DESCRIPTION:

2.1 GENERAL:

We propose a lung sound classification system designed to support medical professionals in the early identification and diagnosis of respiratory conditions. By leveraging advanced machine learning approaches, the system aims to categorize lung sound recordings into distinct types, including normal breath sounds, wheezes, crackles, and other abnormal patterns. To facilitate this, we will apply powerful feature extraction methods—specifically Mel-

Frequency Cepstral Coefficients (MFCCs) and Chroma features—to capture both the spectral and temporal properties of the audio signals. These features will serve as inputs to a carefully designed neural network model, which will be trained using a comprehensive and varied dataset of lung sound recordings. The system's effectiveness will be evaluated based on its classification accuracy, with the goal of providing healthcare practitioners with a reliable tool to enhance diagnostic precision and improve clinical outcomes.

2.2 METHODOLOGIES

2.2.1 MODULES NAME:

1. Data Acquisition

2. Conversational AI Model Development

3. Wizard-of-Oz Technique Implementation

4. Benchmarking and Evaluation

5. User Interface (UI) Integration

1. Feature Extraction Using Mel-Frequency Cepstral Coefficients (MFCC):

- Audio Preprocessing: The raw audio input is initially processed to eliminate background noise and adjust volume levels for consistency.

- MFCC Feature Extraction: Mel-Frequency Cepstral Coefficients are computed from the cleaned audio. These coefficients effectively capture the key elements of the sound spectrum by analyzing the short-term power spectrum, which is especially useful in recognizing the unique features of gunshot sounds.

2. Deep Learning with Neural Networks Incorporating MFCC and Chroma:

- Neural Network with MFCC and Chroma Features: A pre-trained deep learning model is employed to classify audio into multiple categories, including gunshots. It uses both MFCC and Chroma features to improve classification accuracy.

- Improved Detection Accuracy: The model's predictions are used to supplement and verify those of an SVM-based classifier. This hybrid approach leverages the strengths of the deep learning model to enhance the overall sound recognition capabilities.

3. System Integration and Real-Time Operation:

- Real-Time Processing Pipeline: The integrated model is deployed in a real-time audio processing framework capable of identifying gunshot events as they happen. The pipeline handles continuous input, extracts features, classifies the audio, and delivers detection results promptly.

- System Optimization: To achieve fast and efficient processing, the system is fine-tuned using methods like parallel computation and hardware acceleration, ensuring timely and accurate detection.

4. Validation and Evaluation:

- Dataset Development: A diverse and labeled audio dataset containing various types of gunshot and non-gunshot sounds is prepared for training and evaluating the performance of the system.

2.3 TECHNIQUES OR ALGORITHM USED

2.3.1 EXISTING TECHNIQUE:

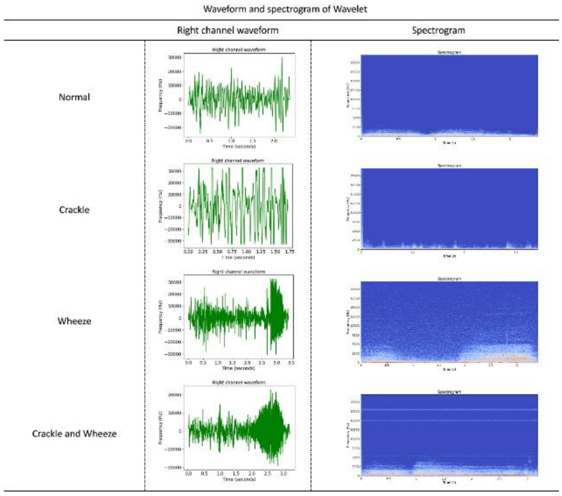
- Residual Connections for Gradient Flow: BiResNet incorporates skip connections—also known as residual connections—that facilitate the smooth flow of gradients during backpropagation. This architecture helps address the vanishing gradient issue, making it possible to train much deeper networks by focusing on learning residuals instead of direct transformations.
- Bidirectional Data Processing: A key feature of BiResNet is its ability to process information in both forward and reverse directions. This bidirectional flow is especially beneficial for sequential or time-series data, where understanding context from both past and future points significantly enhances the model's ability to detect complex relationships and temporal patterns.

- Enhanced Feature Representation: By combining residual learning with bidirectional processing, BiResNet is capable of generating richer and more informative feature representations. This synergy leads to improved performance in tasks that demand a detailed understanding of structured data, such as in medical diagnostics or language modeling.

2.3.2 PROPOSED TECHNIQUE USED OR ALGORITHM USED:

Mel-Frequency Cepstral Coefficients (MFCCs), Chroma features, and neural networks together form an effective framework for analyzing and categorizing audio data. MFCCs transform an audio signal's frequency components into the Mel scale, which closely aligns with how humans perceive pitch. This process captures the detailed spectral characteristics of the sound, making it highly useful in applications like speech and sound classification. In contrast, Chroma features emphasize the harmonic elements of audio by assigning energy to the twelve semitones of the chromatic musical scale. These features are particularly beneficial for music-related tasks, such as identifying musical keys and classifying genres. When both MFCC and Chroma features are used as input to neural networks—systems composed of multiple interconnected layers capable of learning complex data representations—the accuracy of audio classification improves significantly. These networks can learn subtle patterns in auditory signals, enabling more precise identification and analysis in areas such as speech recognition, music analysis, and environmental sound detection.

3 RESULT:



4.FUTURE ENHANCEMENTS:

Future developments in lung sound classification could benefit from the implementation of more sophisticated deep learning models, such as transformers and graph neural networks. These architectures are capable of capturing intricate patterns and long-range dependencies within audio signals, which may result in improved classification accuracy and diagnostic performance.

5 CONCLUSION:

In summary, this study introduces an innovative method for the automated classification of lung sounds using state-of-the-art machine learning techniques. Through the extraction of meaningful audio features and the development of a reliable neural network model, the research highlights the potential of such systems to support medical practitioners in the early identification and diagnosis of respiratory conditions.

6 REFERENCES:

[1] 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. [2] 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.

[3] 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

[4] 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.

[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 learning," Sci. Rep., vol. 12, no. 1, pp. 1–5, Aug. 2022.

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

[7] 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.

[8] 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.

[9] 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.

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

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

[12] 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. [13] 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.

[14] 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.

[15] 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. [16] N.A.M.AriffandA.R.Ismail,"StudyofAdamandadamaxoptimizerson AlexNet architecture for voice biometric authentication system," in Proc. 17th Int. Conf. Ubiquitous Inf. Manage. Commun. (IMCOM), Jan. 2023, pp. 1–4.