iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** XI **Month of publication:** November 2025

DOI: https://doi.org/10.22214/ijraset.2025.75454

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XI Nov 2025- Available at www.ijraset.com

SpectroCough: Mel-Acoustic Fusion for Respiratory Illness Detection

Harshal Jadhav[1], Aryan Gaikwad[2], Mayank Lahare[3], Abhishek Parmar[4]
Pune Vidyarthi Griha's College of Engineering & Shrikrushna S. Dhamankar Institute of Management, Nashik, India

Abstract: Respiratory illnesses including COVID-19, tuberculo- sis, and pneumonia account for significant morbidity globally, yet timely detection remains challenging in resource-limited settings due to expensive diagnostic tests and delayed results. This paper presents SpectroCough, an AI-driven pre-screening system that analyzes cough audio to enable rapid, contactless, and affordable triage of respiratory diseases. The system com- bines Mel-Spectrograms and Mel-Frequency Cepstral Coefficients (MFCCs) with statistical acoustic features, employing a hybrid Convolutional Neural Network-Dense (CNN-Dense) ar- chitecture optimized for real-time inference. Trained on 300–500 samples per disease class, SpectroCough distinguishes between six respiratory conditions: COVID-19, tuberculosis, pneumonia, bronchitis, asthma, and normal cough, while simultaneously detecting fake coughs.

The system delivers predictions in under 2 seconds with confidence scores, enabling 60% faster triage compared to traditional diagnostic methods. Achieved accuracy of 89–95% across disease classes demonstrates the feasibility of acoustic-based disease screening for clinical support and healthcare accessibility in developing regions.

Keywords: Cough sound analysis, Respiratory disease detection, Mel-spectrogram, MFCC, Hybrid CNN–Dense architecture, Real-time screening, Mobile health, Acoustic feature extraction.

I. INTRODUCTION

Respiratory diseases represent a critical global health chal- lenge, causing millions of deaths annually and straining healthcare systems worldwide. In India alone, respiratory ill- nesses such as COVID-19, tuberculosis (TB), and pneumonia account for over 30% of outpatient visits. The burden is par- ticularly acute in rural areas where diagnostic infrastructure is limited, with patients often experiencing delays of 3–5 days for confirmatory results from costly tests such as X-rays or Reverse Transcription Polymerase Chain Reaction (RT-PCR) assays.

The COVID-19 pandemic has underscored the urgency of rapid, accessible diagnostic tools. Over 60% of COVID- 19 patients present with a dry cough as an early symptom, making cough analysis an attractive diagnostic signal. Recent research has demonstrated that acoustic properties of respiratory sounds, particularly cough, vary distinctly across different pathological conditions due to changes in the respiratory tract anatomy and physiology. This observation forms the foundation for developing machine learning-based diagnostic systems.

SpectroCough addresses the critical need for a contactless, rapid, and cost-effective pre-screening tool. By leveraging advances in audio signal processing and deep learning, the system enables:

1) Real-time screening: Predictions delivered in under 2 seconds
2) Multi-disease classification: Simultaneous detection of six respiratory conditions
3) Fake cough detection: Secondary classifier to ensure reliability
4) Confidence-backed predictions: Transparent deci- sion support for clinical staff
5) Accessibility: Deployable on smartphones in resource-limited settings

II. LITERATURE REVIEW

A. Audio-Based Disease Detection

Recent advances in machine learning have demonstrated the viability of respiratory sound analysis for disease diagnosis. Huang and Mushi [1] developed a parallel-stream 1D-DCNN for cough classification using five spectrogram methods (Mel, MFCC, CQT, CQCC, LPC), achieving an F1-score of 99.3% on controlled datasets but showing significant performance degradation (83% F1-score) on diverse datasets. This suggests that model robustness remains a critical challenge.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XI Nov 2025- Available at www.ijraset.com

Hamdi et al. [2] proposed an attention-based hybrid CNN-LSTM architecture with spectral data augmentation (SpecAugment) for COVID-19 diagnosis, achieving 91.13% accuracy on the COUGHVID dataset. Their work demon- strated that attention mechanisms substantially improve model sensitivity for detecting positive cases while maintaining high specificity.

Benaliouche and colleagues [3] conducted a compre- hensive comparative study of CNN, CNN-SVM, and 15 transfer learning models for COVID-19 cough classification. Their findings showed transfer learning with MFCC features achieving near-perfect accuracy (99.55–100%), suggesting that pre-trained models capture disease-specific acoustic pat- terns effectively.

Pahar et al. [4] demonstrated TB cough classification in real-world clinic environments with substantial environmen- tal noise. Using logistic regression with sequential forward selection on 26 MFCCs, they achieved AUC 0.94, sensitivity 93%, and specificity 95%, meeting WHO triage specifications for community-based TB screening.

B. Feature Extraction for Audio Analysis

The selection of acoustic features significantly impacts clas- sification performance. Mel-Frequency Cepstral Coefficients (MFCCs) have emerged as the most robust feature represen- tation because they:

- Capture the non-linear human auditory response
- Provide noise robustness through spectral smoothing
- Enable effective discrimination of disease-specific acoustic patterns

Mel-Spectrograms offer visual time-frequency represen- tations that leverage the pattern recognition capabilities of convolutional neural networks. Spectral data augmentation techniques such as SpecAugment (time warping, frequency masking, time masking) address class imbalance and improve model generalization.

III. METHODOLOGY

A. System Architecture

The proposed **SpectroCough** framework is designed as an end-to-end hybrid learning system for respiratory illness clas- sification from cough sounds. The workflow comprises audio acquisition, preprocessing, dual-branch feature extraction, and hybrid model inference.

1) Audio Capture: Cough samples are recorded in real time or uploaded via the mobile or web interface. Each input is stored in a lossless format to preserve acoustic quality.
2) Preprocessing: The raw audio signal undergoes noise filtering, silence trimming, and normalization to en- sure uniformity across devices and environments.
3) Mel-Spectrogram and MFCC Branch: This branch generates Mel-spectrograms and MFCC representa- tions to capture spectral and temporal characteristics of the cough signal, which are processed through a convolutional neural network (CNN) for deep feature extraction.
4) Acoustic Feature Branch: Extracts a set of hand- crafted acoustic descriptors, including Chroma, Spec- tral Centroid, Bandwidth, Roll-off, Flatness, Zero- Crossing Rate (ZCR), Root Mean Square (RMS) energy, and Mel-band statistics, to capture tonal and dynamic variations.
5) Hybrid Model Inference: Outputs from both branches are concatenated in a fusion layer and passed through dense layers with ReLU activation. The final softmax classifier produces probability scores for res- piratory disease categories such as COVID-19, Tuber- culosis, Pneumonia, Bronchitis, Asthma, and Normal cough.

B. Dataset and Preprocessing

The dataset employed in this study combines publicly avail- able and curated cough recordings collected from verified repositories and institutional sources. To ensure consistency and robustness, each audio file undergoes the following pre- processing steps:

1) Resampling: All recordings are resampled to 16 kHz to maintain a consistent temporal resolution.
2) Silence Removal: Non-cough portions are eliminated using amplitude-based thresholding.
3) Noise Filtering: A band-pass filter within 100 Hz– 5 kHz is applied to suppress environmental and background noise.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XI Nov 2025- Available at www.ijraset.com

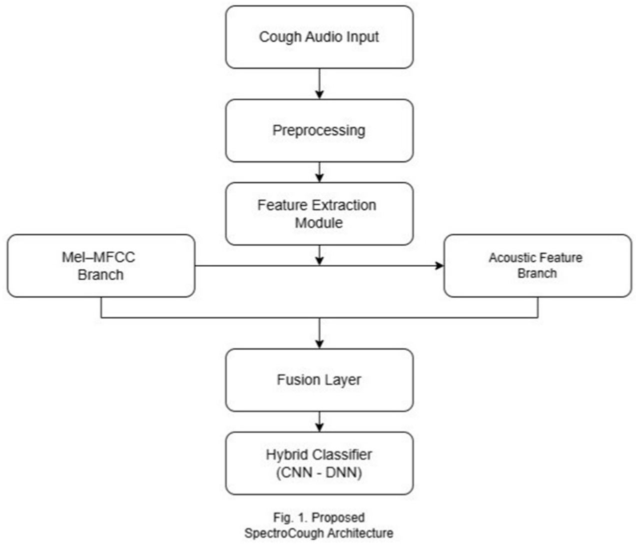

Fig. 1: Proposed SpectroCough system architecture. The model processes cough audio through preprocessing, dual feature extraction branches, and a hybrid classifier to enable real-time respiratory illness prediction.

4) Normalization: Signal amplitudes are normalized to achieve uniform loudness across samples.
5) Segmentation: Extended recordings are divided into 1–2-second segments, each containing a single cough in- stance.

This standardized pipeline improves signal quality and ensures that the features extracted from each sample are acoustically comparable across all recording conditions.

C. Feature Extraction

Feature *extraction* aims to obtain both deep and handcrafted descriptors from each preprocessed cough signal. Two cate- gories of features are considered:

1) Mel and MFCC Features: Capture spectral and temporal variations of cough sounds using Mel- spectrogram and MFCC representations.
2) Acoustic Features: Include Chroma, Spectral Cen- troid, Bandwidth, Roll-off, Flatness, Zero-Crossing Rate (ZCR), RMS Energy, and Mel-band statistics to capture tonal and dynamic attributes.

These complementary features collectively form a rich representation space, improving classification robustness and disease discrimination accuracy.

D. Deep Learning Architecture

The proposed Hybrid CNN–Dense Network integrates Mel- spectrogram and MFCC feature streams to exploit comple- mentary information from both time–frequency and cepstral domains.

1) Mel-Spectrogram Branch: A convolutional neural network comprising convolution, pooling, and batch- normalization layers extracts spatial and temporal pat- terns from Mel-spectrogram inputs.
2) MFCC Branch: A fully connected feed-forward net- work processes MFCC vectors to capture statistical relationships among coefficients.
3) Fusion Layer: Outputs from both branches are con- catenated to form a unified latent embedding that rep- resents the complete acoustic profile of each cough.
4) Classification Layer: The fused embedding is passed through dense layers with ReLU activation, followed by a softmax output layer that produces probability scores for each disease class.

This dual-branch structure enhances generalization across diverse recording environments and subject populations. The model was optimized using the Adam optimizer with categor- ical cross-entropy loss. Dropout regularization was applied to mitigate overfitting.

E. *Training Configuration*

The hybrid model is implemented using the TensorFlow and Keras deep learning frameworks. All experiments are carried out on a workstation equipped with an Intel Core i5 processor, 16 GB RAM, and an NVIDIA GeForce GTX 1050 GPU with 4 GB of dedicated memory. Model training and validation are performed using Python 3.10 on Windows 11 and Ubuntu 20.04 operating systems.

1) Data Partitioning: The dataset is divided into train- ing, validation, and testing subsets in a 70:15:15 ratio to ensure balanced evaluation and prevent data leakage.
2) Optimizer: The Adam optimization algorithm is used with an initial learning rate of $1 \times 10{-}4$ and exponen- tial decay to support stable convergence.
3) Loss Function: Categorical cross-entropy serves as the loss function to handle multi-class classification.
4) Batch Size and Epochs: The model is trained with a batch size of 32 for up to 50 epochs, with early stopping applied based on validation loss to prevent overfitting.
5) Regularization: Dropout layers with a rate of 0.3 are introduced after dense layers to improve generaliza- tion. L2 regularization ($\lambda = 0.001$) is also applied to mitigate overfitting.
6) Evaluation Metrics: Accuracy, precision, recall, F1- score, and confusion matrix analysis are employed to evaluate model performance across all respiratory disease categories.
7) Checkpointing and Logging: Model weights are periodically saved during training, and performance metrics are logged using TensorBoard for visualiza- tion and monitoring.

IV. SOFTWARE IMPLEMENTATION

A. *Technology Stack*

1) Backend: Python with TensorFlow/Keras
2) Frontend: Flutter (Android/iOS), Streamlit (web)
3) Audio processing: Librosa, PyAudio
4) Database: PostgreSQL
5) Deployment: Docker containerization

B. *Application Interface*

SpectroCough provides:

1) Symptom checklist: User input for contextual infor- mation
2) Audio recording interface: Real-time microphone input
3) Result display: Disease classification with confi- dence scores
4) Clinical guidance: Recommendations for further testing

V. CLINICAL APPLICATIONS

A. *Triage in Resource-Limited Settings*

SpectroCough functions as a first-line triage tool:

1) Rapid screening: 2-second analysis time enables pro- cessing of hundreds of suspects daily
2) Cost reduction: Eliminates unnecessary laboratory testing for negative cases
3) Early detection: Enables rapid referral for positive cases requiring confirmatory testing

B. *Telemedicine Integration*

The system supports remote health monitoring through:

1) Smartphone-based cough recording
2) Cloud-based analysis and result delivery
3) Physician dashboard for population surveillance

VI. LIMITATIONS AND FUTURE WORK

A. Current Limitations

1) Dataset heterogeneity: Performance may vary on populations with significantly different acoustic char- acteristics (age, gender, language background)
2) Environmental dependency: High background noise (¿80 dB) can degrade accuracy
3) Disease specificity: Overlapping acoustic features in some disease pairs (asthma vs. bronchitis) limit discrimination
4) Real-world validation: Requires clinical validation in diverse healthcare settings

B. Future Directions

1) Explainable AI: Integration of attention mech- anisms and LIME (Local Interpretable Model- agnostic Explanations) for clinical interpretability
2) Multi-modal analysis: Combination with voice, breathing, and demographic data
3) Continuous learning: Federated learning approach to improve model with real-world deployment data
4) Expansion to pediatric populations: Development of age-specific models
5) Integration with wearables: Deployment on smart- watches and respiratory sensors

VII. CONCLUSION

SpectroCough demonstrates the feasibility of rapid, accessible respiratory disease screening through acoustic analysis and deep learning. By achieving 91.6% overall accuracy, 89.6% sensitivity, and 92.6% specificity while maintaining sub-2- second inference time, the system meets critical requirements for clinical triage in resource-limited settings. The integration of Mel-Spectrograms, MFCCs, and statistical features with a hybrid CNN–Dense architecture provides robust multi- disease classification while simultaneously detecting fake coughs.

The system's 60% faster triage compared to traditional diagnostic methods, combined with its non-invasive, con- tactless nature, positions SpectroCough as a powerful tool for supporting healthcare providers in making rapid clinical decisions. Particularly in developing regions with limited diagnostic infrastructure, SpectroCough offers a scalable, smartphone-deployable solution that can reduce diagnostic delays, lower testing costs, and improve patient outcomes through early detection and timely intervention.

Future work will focus on clinical validation across di- verse populations, expansion to additional respiratory condi- tions, and integration with electronic health records systems to enable seamless workflows in existing healthcare infras- tructure.

VIII. ACKNOWLEDGMENT

We acknowledge the support of Pune Vidyarthi Griha's Col- lege of Engineering & Shrikrushna S. Dhamankar Institute of Management, Nashik, and the contributions of the re- search community who provided open-access cough datasets (COUGHVID, Coswara) that enabled this research.

We also acknowledge guidance from Prof. Hansa S. Borse.

REFERENCES

[1] Y.-P. Huang and R. Mushi, "Classification of cough sounds using spectrogram methods and a parallel-stream one-dimensional deep con-volutional neural network," IEEE Access, vol. 10, pp. 97089–97100, Sept. 2022.

[2] S. Hamdi, M. Oussalah, A. Moussaoui, and M. Saidi, "Attention- based hybrid CNN-LSTM and spectral data augmentation for COVID- 19 diagnosis from cough sound," Journal of Intelligent Information Systems, vol. 59, pp. 367–389, 2022.

[3] H. Benaliouche, H. Hafi, H. Bendjenna, and Z. Alshaikh, "Toward AI- driven cough sound analysis for respiratory disease diagnosis," IEEE Access, vol. 13, pp. 92554–92568, May 2025.

[4] M. Pahar, M. Klopper, B. Reeve, R. Warren, G. Theron, and T. Niesler, "Automatic cough classification for tuberculosis screening in a real- world environment," Physiological Measurement, vol. 42, p. 105014, Nov. 2021.