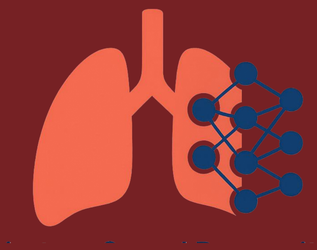
Respiratory Sound Recognition Using Deep Learning: A Comparative Study of CNN, ResNet50, and DenseNet121 Architectures

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Author: Giagkos Stylianos

Affiliation: MSc in Data Science, Athens University of Economics and Business (AUEB)

Supervisor: Professor Themos Stafylakis



ABSTRACT

This study explores the application of deep learning models—Custom Convolutional Neural Networks (CNN), ResNet50, and DenseNet121—for the classification of respiratory sounds into ten diagnostic categories. A rigorous preprocessing pipeline transforms raw lung sound recordings into 2D log-mel spectrograms using DFT-based filtering, temporal segmentation, and augmentation techniques. Models are evaluated using patient-wise splits, weighted loss functions, and macro F1-scores to account for class imbalance. Results demonstrate that transfer learning with DenseNet121 and ResNet50 offers superior generalization compared to baseline CNNs, with DenseNet121 achieving the highest ROC AUC (0.81) and ResNet50 delivering the best balanced accuracy (74%). Despite progress, rare class prediction remains limited. This study highlights the clinical potential of spectrogram-based deep learning for automated auscultation.

INTRODUCTION

Auscultation is fundamental in diagnosing respiratory conditions, but it is subjective and varies across clinicians. Automated classification of lung sounds offers an objective, scalable alternative—especially valuable in remote or low-resource environments. Pathological sounds like wheezes, crackles, and rhonchi correlate with diseases such as asthma, COPD, and pneumonia. Yet, automated analysis is challenged by non-stationary signals, noise, and class imbalance. Deep learning, particularly CNNs applied to spectrograms, has shown promise in audio-based classification. This study evaluates both custom CNNs and pre-trained models (ResNet50, DenseNet121) using a unified pipeline to assess generalization, robustness, and clinical applicability.

METHODOLOGY

Dataset & Preprocessing

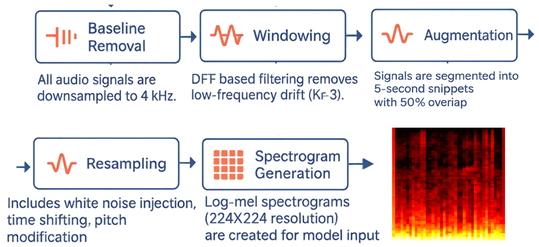
Resampling: All audio signals are downsampled to 4kHz.

Baseline Removal: DFT-based filtering removes low-frequency drift (K=8).

Windowing: Signals are segmented into 5-second snippets with 50% overlap.

Augmentation: Includes white noise injection, time shifting, pitch modification.

Spectrogram Generation: Log-mel spectrograms (224 × 224 resolution) are created for model input.



Models

CustomCNN: 3-block convolutional architecture trained from scratch.

ResNet50: Pretrained on ImageNet, adapted for grayscale inputs, with a residual classifier head.

DenseNet121: Pretrained with dense connectivity, enabling feature reuse and improved flow.

CORAL (optional): Domain adaptation method minimizing distributional mismatch via covariance alignment.

Training & Evaluation

Loss Function: Weighted CrossEntropyLoss (class-frequency based)

Optimizers: AdamW or SGD (depending on experiment)

Schedulers: ReduceLROnPlateau for dynamic learning rate tuning

Metrics: Accuracy, Macro F1-score, ROC AUC, and confusion matrices

Early Stopping: Triggered after 6 epochs of non-improvement

RESULTS

Experiment 1 (Baseline Training – AdamW)

Model	Accuracy	Macro F1	ROC AUC
CustomCNN	59%	0.56	0.32
ResNet50	71%	0.62	0.28
DenseNet121	72%	0.59	0.33

- ResNet50 and DenseNet121 outperform CustomCNN
- Poor discrimination (AUC < 0.35) despite high accuracy
- Rare class prediction (e.g., Lung Fibrosis, Pleural Effusion) fails entirely

Experiment 2 (SGD for CNN, Regularization Tweaks)

Model	Accuracy	Macro F1	ROC AUC
CustomCNN	52%	0.46	0.54
ResNet50	**74%**	**0.67**	**0.57**
DenseNet121	66%	0.61	0.32

- ResNet50 and DenseNet121 outperform CustomCNN
- Poor discrimination (AUC < 0.35) despite high accuracy
- Rare class prediction (e.g., Lung Fibrosis, Pleural Effusion) fails entirely

Experiment 3 (CORAL – Domain Adaptation)

Model	Accuracy	Macro F1	ROC AUC
CustomCNN	61%	0.55	0.28
ResNet50	68%	0.52	0.37
DenseNet121	67%	0.41	0.81

- ResNet50 and DenseNet121 outperform CustomCNN
- Poor discrimination (AUC < 0.35) despite high accuracy
- Rare class prediction (e.g., Lung Fibrosis, Pleural Effusion) fails entirely

RECOMMENDATIONS

Technical Improvements

Focal Loss or Class-balanced Loss for minority class emphasis

Synthetic Data Generation (GANs or SMOTE) for rare conditions

Threshold Calibration (e.g., temperature scaling) to improve probabilistic outputs

Adversarial Domain Adaptation (DANN) as a more dynamic alternative to CORAL

Clinical Integration

Model **compression and quantization** for mobile deployment

Use of **interpretable features** for clinician trust

Validation on prospective real-world datasets

CONCLUSION

This study benchmarks three CNN-based architectures for the automated classification of respiratory sounds using spectrograms. **ResNet50 with AdamW** provided the best all-around performance, while **DenseNet121 with CORAL** showed exceptional class separability (AUC = 0.81) but lower class-level balance. The analysis revealed persistent challenges with **imbalanced data** and **acoustically overlapping conditions**. Future directions include smarter data balancing, domain adaptation, and real-world validation. The research supports the **feasibility of deploying deep learning for lung auscultation**, especially when paired with transfer learning and robust preprocessing.

ACKNOWLEDGEMENTS

This research was conducted as part of the MSc in Data Science program at Athens University of Economics and Business (AUEB).Special thanks to Professor Themos Stafylakis for supervision and guidance.

We acknowledge the use of OpenAI's ChatGPT for editorial, structural, and technical assistance throughout the preparation of this research project.