World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(RESEARCH ARTICLE)

Multi-Class Lung Disease Classification Using Google's HeAR Foundation Model and MFCC Features: A Comparative Study of Performance and Data Efficiency

Brighton Mukundwi [1,*] and Delvin Tadiwa Vengesai [2]

[1] *Department of Data Analytics and Visualisation, Faculty of Computer Science, Yeshiva University; ORCiD: 0009-0003-8516-9656*
[2] *Department of Biological Sciences and Ecology, Faculty of Science, University of Zimbabwe; ORCiD: 0009-0001-4948-5729*

World Journal of Advanced Research and Reviews, 2026, 30(02), 1902-1913

Publication history: Received on 10 April 2026; revised on 20 May 2026; accepted on 22 May 2026

Article DOI: https://doi.org/10.30574/wjarr.2026.30.2.1459

Abstract

Over 454 million individuals worldwide suffer from chronic respiratory conditions, with low- and middle-income countries bearing a disproportionate burden due to inadequate diagnostic facilities. Although foundation audio models remain underexplored in this field, automated classification of respiratory sounds could support early disease detection at scale. This study evaluates Google's Health Acoustic Representations (HeAR) model for multi-class lung disease classification and directly compares it with conventional Mel-frequency cepstral coefficient (MFCC) features across five classifiers, including SVM, Gradient Boosting, and MLP. Using the Asthma Detection Dataset Version 2, audio was segmented into 3,602 two-second clips, balanced with SMOTE, and evaluated on a stratified held-out test set. Model performance was assessed using training data fractions ranging from 10% to 100% in a data efficiency experiment. The best HeAR-based model, MLP, achieved a macro F1-score of 84.5% and accuracy of 86.4%, while MFCC features combined with Gradient Boosting produced the highest overall performance, with 88% accuracy and 87% macro F1-score. HeAR embeddings consistently outperformed linear classifiers under limited data conditions. At 10% training data, HeAR SVM achieved a macro F1-score of 70%, compared to 55.8% for MFCC SVM. While MFCC features with non-linear ensembles delivered superior peak performance on controlled single-source data, HeAR embeddings produced a more linearly separable feature space, enabling stable classification with less labelled data, making them suitable for resource-limited clinical settings.

Keywords: Health Acoustic Representations; MFCC Features; Respiratory Sound Analysis; Digital Health; Low-Resource Settings

1. Introduction

One of the most important global public health issues of the twenty-first century is chronic respiratory disorders. They caused over 4 million fatalities and impacted over 454 million people globally in 2019, with COPD and asthma together accounting for the majority of this burden (1). The distribution of this load is not uniform. Due in part to restricted access to specialised clinical treatment and diagnostic facilities, low and middle-income countries (LMICs) face a disproportionate amount of respiratory illness mortality and disability-adjusted life years (2). Timely intervention in these circumstances depends on the early and precise characterisation of respiratory diseases. In clinical practice, pulmonary auscultation is still the major non-invasive method of diagnosing respiratory diseases. However, manual lung sound interpretation is intrinsically subjective, heavily reliant on clinician expertise, and prone to inter-observer variability; interns' and residents' classification accuracy can be as low as 53–69% (3). Growing interest in automated, AI-powered pulmonary sound classification systems that can offer objective, scalable, and repeatable diagnostic help has resulted from these limitations.

* Corresponding author: Brighton Mukundwi

Recent developments in deep learning and machine learning have shown excellent results on challenges involving the classification of respiratory sounds. Transformer-based models and CNN-LSTM hybrids are examples of complex deep learning architectures that are still widely used. Although these models perform quite well on benchmarks, they usually demand extensive computational resources and large labelled datasets, both of which are difficult to come by in clinical settings with limited resources. The use of foundation audio models, large pre-trained models that produce generalised acoustic embeddings, in combination with lightweight classical classifiers is a complementary and understudied approach. This method requires far less labelled training data and provides a computationally accessible, possibly high-performing alternative. The best foundation model currently available for health acoustic tasks is Google's Health Acoustic Representations (HeAR) model, which was pre-trained on more than 300 million de-identified health-related audio samples. Its use in the classification of respiratory sounds has been minimal thus far. HeAR was utilised in a recent study to detect respiratory diseases, namely binary paediatric asthma, with over 91% accuracy on the SPRSound dataset (4). HeAR has not been used for multi-class lung disease classification in any previous work, nor has its performance been rigorously evaluated with conventional MFCC features under various data efficiency settings.

This paper aims to contribute to that gap. Using an open-source dataset, we demonstrate the application of the HeAR foundation model to multi-class lung disease classification across four clinically diverse conditions: asthma, bronchitis, COPD, and pneumonia, including the "Healthy" class. Five classical classifiers are used to directly benchmark HeAR embeddings against MFCC features. In order to empirically evaluate the hypothesis that foundation model embeddings offer a more stable and data-efficient feature space than handcrafted features, we will carry out a data efficiency experiment in which we train each model on decreasing data fractions. The results directly affect the development of deployable respiratory screening instruments in low-resource environments without access to sizable annotated datasets.

2. Related Work

2.1. Automated Respiratory Sound Classification

Over the past ten years, there has been a significant increase in the automated analysis of respiratory sounds for disease classification, mostly due to the ICBHI 2017 Respiratory Sound Database. Machine learning models can successfully classify abnormal lung sounds, with accuracy ranging from 69.4% to 99.6% for disease classification, according to a systematic review of 62 studies. However, comparability across studies is limited due to inconsistent methodologies and a high risk of bias (5). Deep learning techniques have dominated the field. With dual-channel CNN-LSTM models showing accuracy above 99% on the ICBHI dataset, CNN-LSTM hybrid architectures have emerged as the de facto standard for temporal audio classification tasks (6). Similar impressive results have been shown by multi-feature fusion techniques that combine MFCCs, Mel spectrograms, and chromagrams. CNN-based classifiers achieved 91% accuracy on 10-class lung sound datasets (7). These architectures have a common drawback despite their high stated performance: they require large, well-labelled training datasets, are computationally intensive, and are challenging to implement on edge devices. Noise sensitivity, device variability, and poor model interpretability are recurring obstacles to real-world clinical implementation, according to an assessment of deep learning techniques for lung sound analysis (8).

2.2. MFCC Features in Respiratory Classification

The most popular handmade characteristic for respiratory sound analysis is still mel-frequency cepstral coefficients (MFCC). Research has demonstrated that MFCC features by themselves carry enough acoustic information to categorise five or more kinds of respiratory diseases, with test accuracies of 87–88% possible on limited datasets (9). MFCC-based models have identified respiratory diseases with up to 95.1% accuracy on independent test sets when paired with CNN architectures (10). When assessing the contribution of more sophisticated representation techniques, MFCCs are a suitable and essential benchmark due to their computational simplicity, interpretability, and robust performance on controlled datasets.

2.3. Foundation Audio Models for Health Classification

A novel approach to health acoustic classification has been made possible by the development of large-scale pre-trained audio models. These models encode extensive acoustic knowledge from large datasets and transfer it to downstream tasks with little additional supervised input, as opposed to learning representations from scratch. Strong label efficiency across medical classification tasks has been shown by transfer learning from pre-trained audio models, especially when downstream labelled data is scarce (3). According to a recent study, adapting audio foundation models to cardiac sound classification through ongoing domain-specific pre-training enhanced downstream performance by up to 13% (11). This suggests that domain adaptation of foundation models is still an active and fruitful direction. The most complete

health-specific audio foundation model on the market is Google's HeAR model. It creates 512-dimensional embeddings that capture latent acoustic characteristics pertinent to respiratory health after being pre-trained on more than 300 million de-identified health-related audio recordings, including 100 million cough sounds. The binary paediatric asthma identification study used HeAR embeddings to the SPRSound dataset and achieved over 91% accuracy using SVM, Random Forest, and MLP classifiers, is its first documented use to respiratory disease classification (4). To our knowledge, neither the expansion to multi-class disease categorisation nor the methodical assessment of HeAR's data efficiency characteristics in contrast to conventional features have been investigated. That gap is directly addressed in this work.

2.4. AI-Based Respiratory Screening in Low-Resource Settings

A particular set of properties is required for the design of deployable respiratory screening systems in low-resource environments. This includes low processing overhead, good classification accuracy in noisy, diverse environments, and minimal labelled data requirements. The viability of lightweight transfer learning techniques for COPD and asthma classification with fewer than 350 recordings was shown in a study on cough-based pulmonary disease screening in low-resource settings (12). To guarantee accessibility at the point of service, AI-powered solutions created for LMIC healthcare contexts must give equal weight to computational efficiency and diagnostic accuracy (13). These limitations are taken into consideration when designing the pipeline suggested in this paper, which combines HeAR embeddings with lightweight classical classifiers. The data efficiency experiment assesses the pipeline's applicability for environments with limited annotated respiratory sounds.

3. Methodology

The development of an AI-enabled multi-class lung disease classifier using the Asthma Detection Version 2 dataset, HeAR model, and MFCC Features is described in detail in this section. The development architecture is summarised in Figure 1.

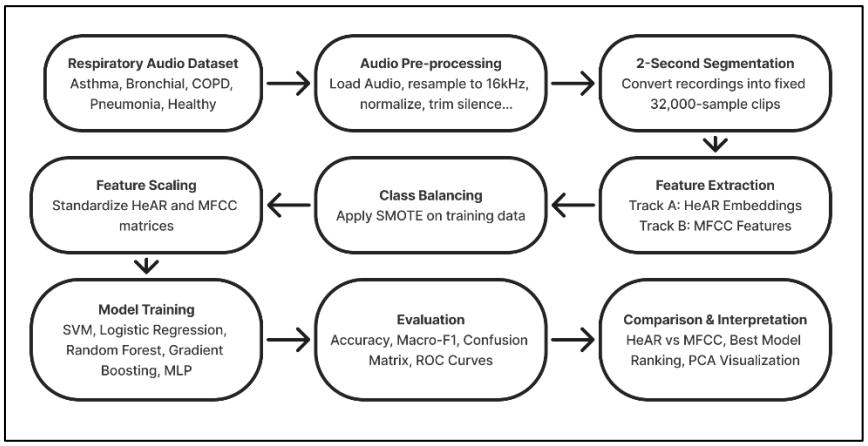

Figure 1 System Architecture for AI-based multi-class lung disease classification using Google's HeAR foundation model and MFCC features

3.1. Dataset

This study uses the publicly available Asthma Detection Dataset Version 2 (14), comprising 1,211 respiratory audio recordings across five clinically labelled classes:

- COPD - 401
- Asthma - 288
- Pneumonia - 285
- Healthy - 133
- Bronchitis - 104

The class labelled "Bronchial" in the source dataset refers specifically to Bronchitis, as confirmed by the original dataset paper (14). All recordings are in *.WAV* format. Mean recording duration was 5.9 seconds, with the majority of recordings at exactly 6 seconds.

3.1.1. Step 1: Audio Pre-processing and Segmentation

All recordings were resampled to 16 kHz mono and amplitude-normalised using peak normalisation. Leading and trailing silence was removed using a 20 dB threshold. Each recording was then segmented into non-overlapping 2-second clips of exactly 32,000 samples, as required by the HeAR model input specification. Clips shorter than 2 seconds were zero-padded. This produced 3,602 clips in total.

- COPD - 1,202
- Pneumonia - 853
- Asthma - 839
- Healthy - 399
- Bronchitis - 309

3.1.2. Step 2: Feature Extraction

Two parallel feature extraction tracks were implemented on the same clip set to enable direct comparison.

Track A - HeAR Embeddings

Each 2-second clip was passed through Google's Health Acoustic Representations (HeAR) model (15), a foundation model pre-trained on over 300 million de-identified health-related audio samples. The model was accessed via the Hugging Face Hub and loaded as a Tensor Flow *savedmodel* layer. Each clip was encoded into a 512-dimensional embedding vector capturing latent acoustic and spectral properties relevant to respiratory health.

Track B - MFCC Features

From the same clips, 40 Mel-frequency cepstral coefficients (MFCCs) were extracted using librosa (16), along with their first- and second-order temporal derivatives (delta and delta-delta). The mean of each coefficient across the time axis was computed, yielding a fixed 120-dimensional feature vector per clip.

3.1.3. Step 3: Dataset Splitting and Class Imbalance Handling

The full clip set was split into training (80%) and test (20%) partitions using stratified random sampling, producing 2,881 training clips and 721 test clips. Stratification preserved class proportions in both splits. SMOTE (17) was applied independently to both feature tracks using k=5 nearest neighbours, exclusively on training data. The test set remained unmodified. After oversampling, all five classes were represented equally at 962 samples each, for 4,810 training samples in total. Features were standardised using zero-mean unit-variance scaling, with parameters fitted on training data only and applied to both splits.

3.1.4. Step 4: Classification Models

Five classifiers were trained on each feature track, producing ten models in total. All were implemented using scikit-learn (18). The classifiers are outlined in Table 1

Table 1 Machine Learning Classifiers used in this study

Model	**Description**
SVM (Linear)	Support Vector Machine with linear kernel, probability estimates enabled.
Logistic Regression	Probabilistic linear classifier, L2 regularisation, maximum 1,000 iterations.
Random Forest	Ensemble of decision trees, 200 estimators
Gradient Boosting	Boosted tree-based ensemble, 150 estimators
MLP Classifier	Multi-layer perceptron neural network. Two hidden layers of 128 and 64 units, ReLU activation.

3.1.5. Step 5: Evaluation

All models were evaluated on the held-out test set. Primary metrics were macro-averaged precision, recall, and F1-score, chosen to weight performance equally across all five classes regardless of class size. Accuracy is reported as a

secondary metric. Confusion matrices and one-vs-rest ROC curves with per-class AUC scores were generated for the best-performing model on each track.

3.1.6. Step 6: Data Efficiency Experiment

To evaluate HeAR performance under data-scarce conditions which are relevant to low-resource clinical deployment, each classifier was retrained on stratified subsets of the training data corresponding to six fractions: 10%, 20%, 30%, 50%, 75%, and 100%. At each fraction, training was repeated across three random seeds (42, 123, and 7) and macro F1 scores were averaged to reduce sampling variance. This was applied to both feature tracks, enabling direct comparison of how HeAR embeddings and MFCC features degrade as labelled training data decreases.

4. Results

4.1. Overall Classification Performance

Table 2 presents macro-averaged precision, recall, F1-score, and accuracy for all ten models on the held-out test set. MFCC + Gradient Boosting achieved the strongest overall performance, followed by MFCC + Random Forest. Among HeAR-based models, MLP performed best. Linear models (SVM and Logistic Regression) performed markedly better on HeAR embeddings than on MFCC features, a pattern discussed in Section 5.

Table 2 Classification Performance of All Models on the Held-Out Test Set

	Track	Model	Accuracy	Macro F1	Macro Precision	Macro Recall
1	MFCC	Gradient Boosting	0.883495	0.872444	0.874586	0.872306
2	MFCC	Random Forest	0.871012	0.856584	0.873641	0.851321
3	HeAR	MLP Classifier	0.864078	0.844809	0.853553	0.838339
4	HeAR	SVM Linear	0.833564	0.821471	0.821851	0.823339
5	HeAR	Logistic Regression	0.828017	0.809992	0.810150	0.810248
6	MFCC	MLP Classifier	0.805825	0.775864	0.790464	0.765636
7	HeAR	Gradient Boosting	0.783634	0.751159	0.760827	0.744222
8	HeAR	Random Forest	0.768377	0.731242	0.759407	0.715279
9	MFCC	SVM Linear	0.722607	0.666911	0.671839	0.669199
10	MFCC	Logistic Regression	0.711512	0.659481	0.656447	0.666062

Macro-averaged metrics weight each class equally regardless of support size. Accuracy is reported as a secondary metric.

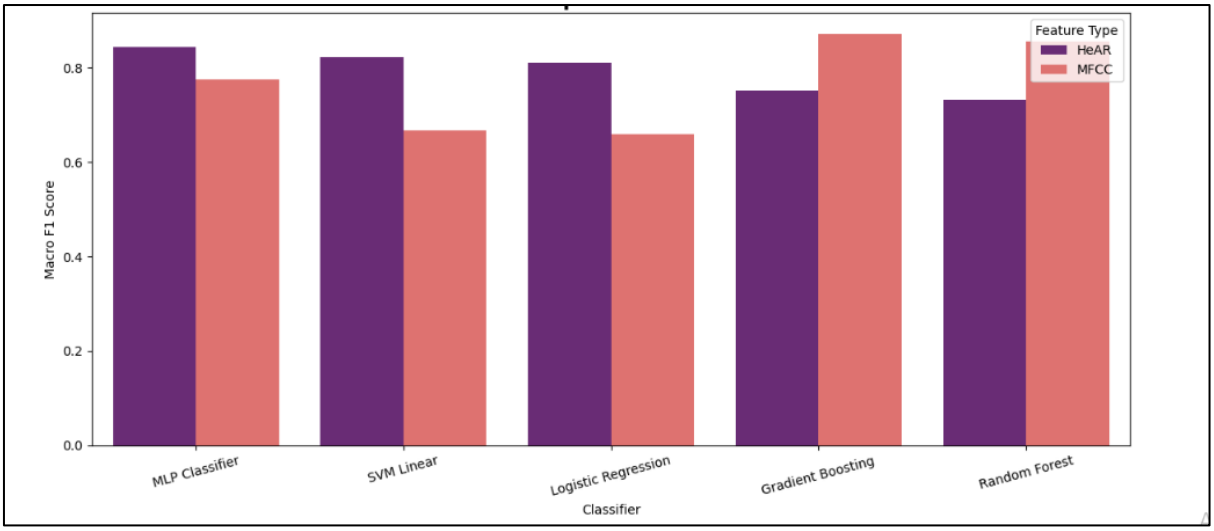

Figure 2 Macro F1 Comparison Across All Models, grouped bar chart, HeAR vs MFCC

4.2. Confusion Matrix Analysis

Figure 3 Confusion Matrix for best HeAR Model (MLP Classifier)

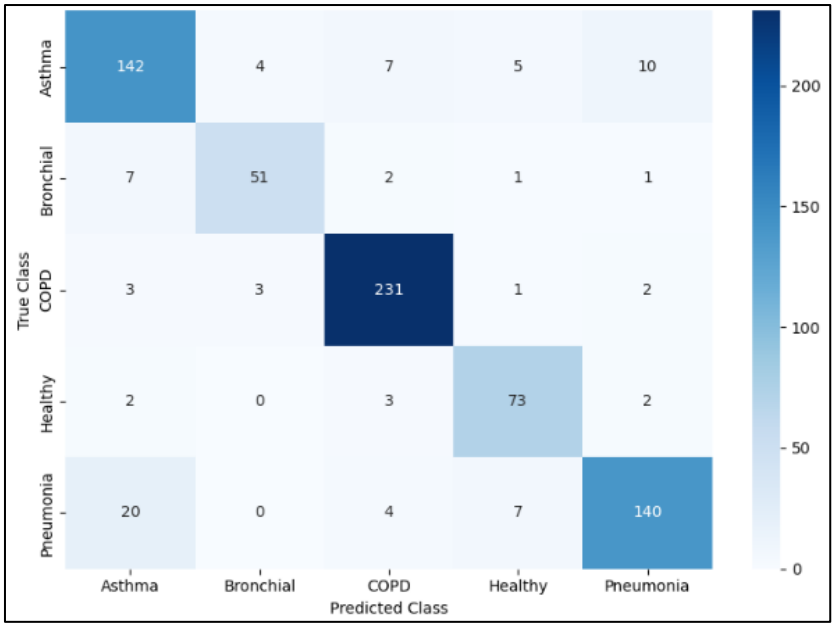

Figure 4 Confusion Matrix for best MFCC Model (Gradient Boosting)

4.3. ROC Curve Analysis

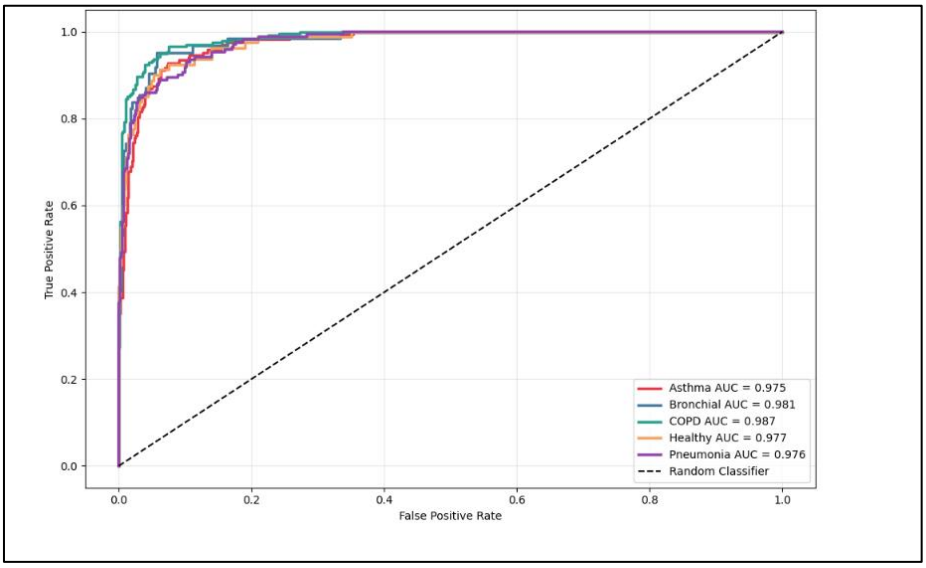

Figure 5 Multi-Class ROC Curve for the best HeAR Model (MLP Classifier)

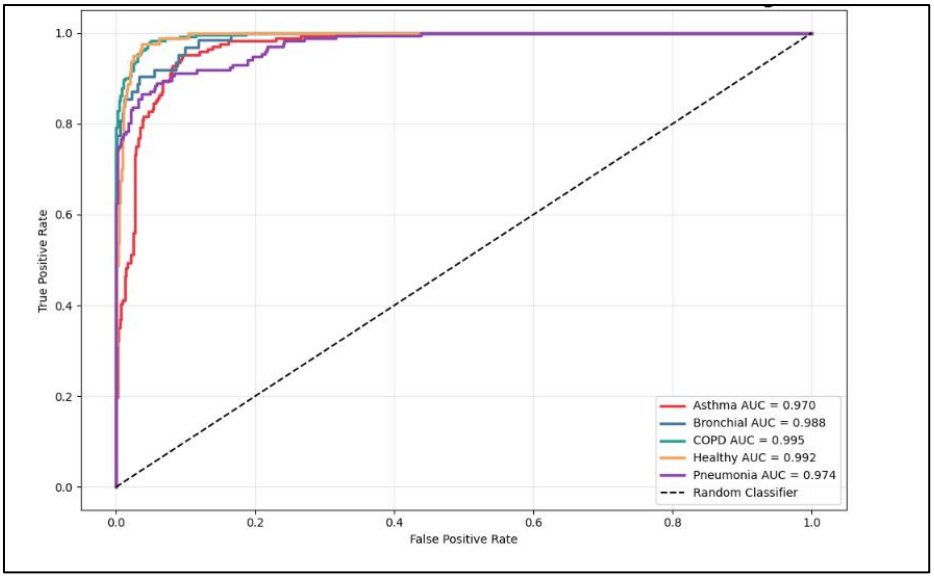

Figure 6 Multi-Class ROC Curve for the best MFCC Model (Gradient Boosting)

4.4. Embedding Space Visualization

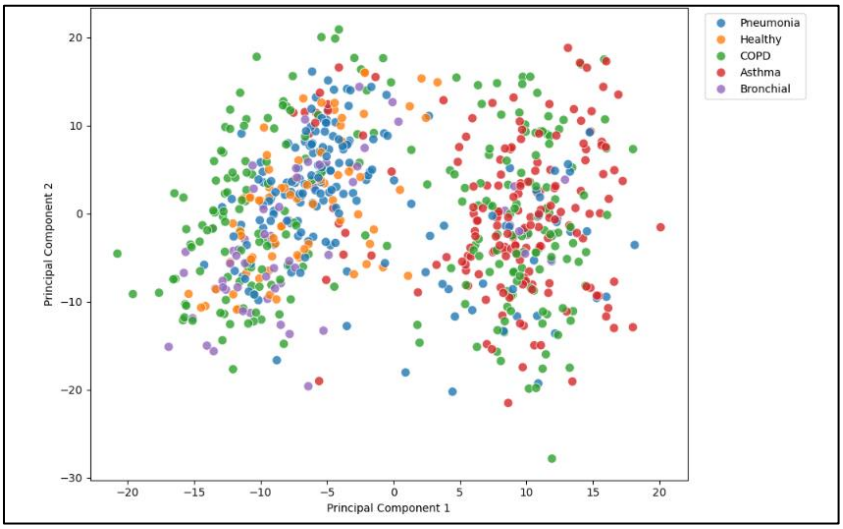

Figure 7 PCA Projection of HeAR Embeddings

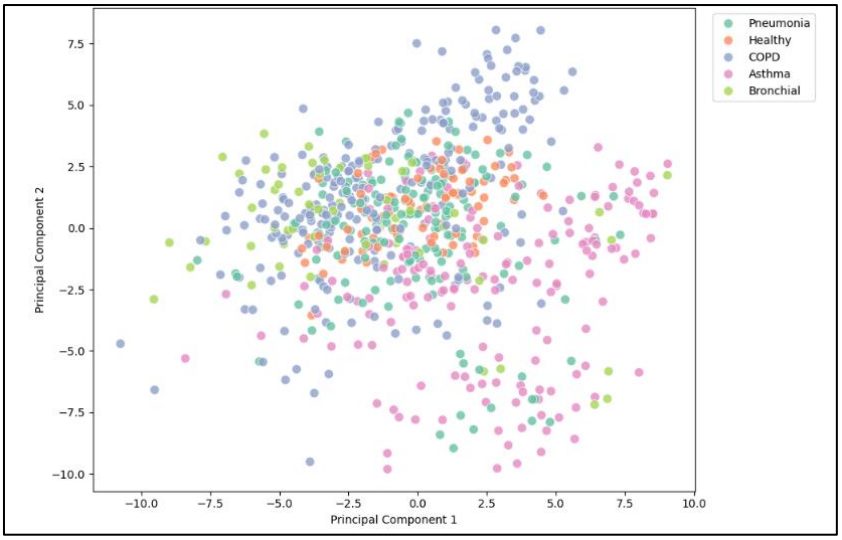

Figure 8 PCA Projection of MFCC Features

4.5. Data Efficiency

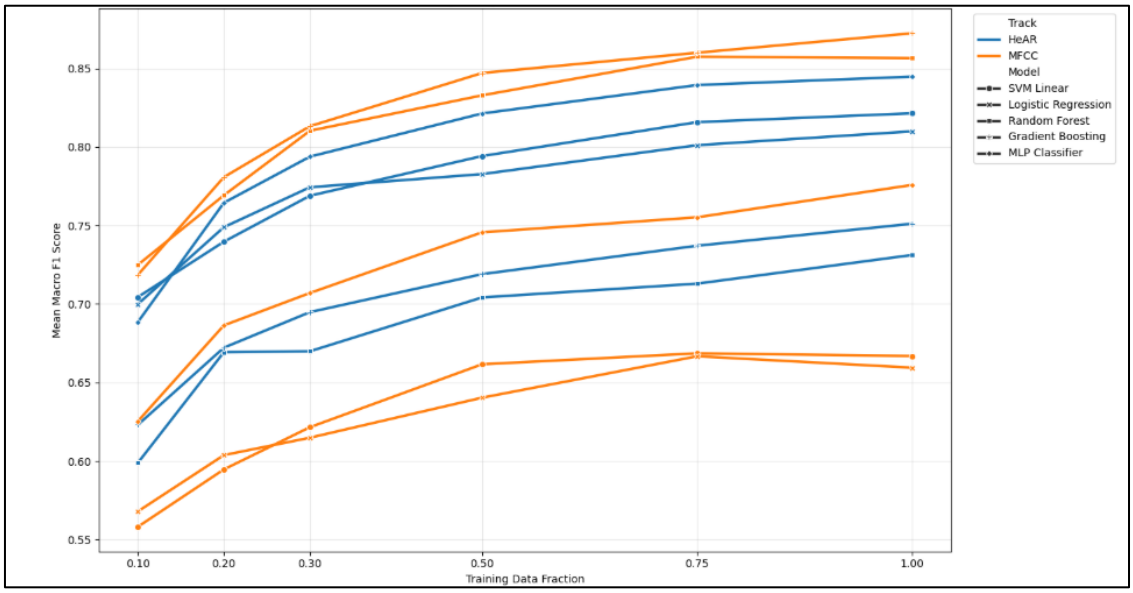

Figure 9 Data Efficiency experiment results (HeAR vs. MFCC)

5. Discussion

5.1. Interpretation of Results

The primary finding is that on peak classification performance, MFCC features combined with Gradient Boosting performed better than the best HeAR-based model, MLP. By embedding richer, more generalizable acoustic representation, foundation models are typically thought to perform better than handcrafted features; however, this is not the case in this investigation. The characteristics of this dataset directly explain the result. The 1,211 recordings are all from the same source, and they are all almost identical in length and acoustically controlled. Here, the fundamental benefit of HeAR, which is the robust generalisation across diverse devices, populations, and noisy environments, is not adequately exploited. Traditional features such as MFCCs can capture disease-relevant spectral patterns with high fidelity when the acoustic environment is clean and consistent, and tree-based ensemble approaches are well-suited to take advantage of the non-linear feature interactions in the 120-dimensional MFCC space. This is in line with more general evidence that on structured, single-source datasets, and gradient boosted decision trees continue to outperform trained representations (19). This finding should not be generalised. HeAR's generalisation attributes would be

expected to close or reverse this difference in a real-world deployment context with diverse recording devices and changeable acoustic conditions, the exact situation in low-resource healthcare. However, this study does not test that assertion.

5.2. Data Efficiency and Linear Separability

HeAR embeddings consistently shown a clinically significant advantage in data-scarce settings for linear classifiers, despite poorer peak performance. At 10% training data, HeAR SVM beat MFCC SVM by 0.146 macro F1, and this difference persisted for all fractions. More linearly separable representations are encoded by foundation models pre-trained on large, diverse datasets, enabling efficient classification with minimal extra supervised data. Strong label efficiency across medical classification tasks has been shown by transfer learning from pre-trained models, especially when downstream labelled data is scarce (3). The data efficiency argument has been supported by recent work on pathological foundation models, which has demonstrated that linear probing on pre-trained representations can match or surpass refined traditional models with significantly less labelled data (20). The complimentary pattern is important. HeAR embeddings enable efficient linear classification under data restrictions, whereas MFCC features need non-linear classifiers and bigger training sets to achieve competitive performance. These show various points of operation on the trade-off between performance, data, and availability.

5.3. Clinical Implications for Low-Resource Settings

It is difficult to overestimate the usefulness of data efficiency when it comes to respiratory illness screening in environments with limited resources. Large respiratory audio datasets must be gathered and labelled using clinical knowledge, recording equipment, and physician time, all of which are limited resources in low-resource environments. Due to the lack of qualified clinical personnel and diagnostic tools like spirometry, pulmonary illnesses are often misdiagnosed or underdiagnosed in such settings (12). A deployable route for community-level screening is provided by a pipeline that uses just 10% of training data to obtain macro F1 of 0.70 without the need for sophisticated deep learning architectures or GPU resources. In addition to diagnosis accuracy, AI-powered systems intended for low-resource clinical deployment must prioritise computational efficiency and accessibility (13). The HeAR + linear classifier combination meets both needs in a manner that complicated architectures simply do not. Context determines the optimal deployment strategy. MFCC + Gradient Boosting provides the best classification results when there is enough labelled data and processing power. HeAR embeddings with lightweight linear classifiers provide better and more consistent performance in situations where labelled data is limited.

5.4. Misclassification Analysis

The optimal HeAR and MFCC Features models had error rates of 13.6% and 11.65%, respectively, which were concentrated near acoustically confusing class boundaries. The boundaries between bronchitis and COPD and asthma and pneumonia showed the most common misclassification patterns. It is well recognised that these separations present auditory difficulties. In brief audio segments, pneumonia and asthma-related sounds occupy partially overlapping frequency bands (21), crackles linked to bronchitis are also clinically associated with pneumonia (22), and COPD and pneumonia share common adventitious lung sounds (23). These results support the suggestion that this pipeline not be used as a stand-alone diagnosis system, but rather as a screening tool to assist clinical triage.

5.5. Limitations

There are several limitations to be aware of. First off, the acoustically homogeneous recordings in the dataset originate from a single cleaned and carefully curated source. This decreases the cross-device generalisation advantage of HeAR directly and probably explains why the peak performance of MFCC features was higher. The results shown here should not be interpreted as proof that MFCC features are typically better than HeAR embeddings for the classification of respiratory diseases; rather, they show how both methods perform in situations that favour handcrafted features. Second, the HeAR model's representational ability on this dataset might have been limited because it was employed as a fixed feature extractor without domain-specific fine-tuning. Third, the acoustic diversity of actual patient recordings may not be entirely replicated by SMOTE-generated synthetic samples for the Bronchitis class. Lastly, findings about generalisability to different demographics and recording environments were limited by the lack of external clinical validation on independent, multi-site cohorts.

5.6. Future Directions

Future research should look toward fine-tuning HeAR using respiratory data particular to a certain domain. Compared to linear probing alone, downstream classification performance has increased by up to 13% when audio foundation models are modified through ongoing domain-specific pre-training (11). The generalisation benefit claimed to HeAR

would be explicitly tested by cross-device validation among various stethoscope types. Important next steps toward deployable clinical tools include multi-label classification to account for co-occurring disorders and extension to bigger, more varied datasets, such as recordings from primary care settings in sub-Saharan Africa and other LMICs.

6. Conclusion

In this work, classical MFCC features and Google's HeAR foundation model embeddings were compared for multi-class lung illness classification under five different conditions: Pneumonia, COPD, bronchitis, asthma, and healthy. The top HeAR-based model, MLP, was surpassed by MFCC features with Gradient Boosting on peak performance at full training data. This is not a universal shortcoming of foundation model embeddings, but rather the strength of tree-based ensemble approaches on organised, acoustically consistent feature spaces. Conversely, the data efficiency experiment revealed HeAR embeddings outperformed MFCC-based linear models by up to 0.146 macro F1 points at 10% training data, a difference that persisted across all data fractions, and they consistently provided superior linear classification at low data volumes. Less tagged data is needed to fully utilise the richer, more linearly separable auditory representation encoded by HeAR embeddings. In practice, MFCC features with gradient boosting provide the best classification accuracy when there is enough labelled data and computational power. Large annotated respiratory audio files cannot be practically gathered in low-resource environments. A deployable, data-efficient substitute that sacrifices a little peak performance for significantly lower data requirements is provided by HeAR embeddings with lightweight linear classifiers. The deployment situation should determine which of these complementary approaches to use. Fine-tuning should be the main focus of future development. To advance this pipeline toward clinical implementation, HeAR on domain-specific respiratory audio, cross-device validation, and extension to bigger and more geographically diverse datasets, including recordings from primary care settings in LMICs.

Compliance with ethical standards

Disclosure of conflict of interest

The authors declare no conflicts of interest.

Statement of ethical approval

Only de-identified audio data that was accessible to the public was used in this investigation. Neither fresh data gathering nor the recruitment of human volunteers took place. Therefore, ethical approval was not necessary. The published terms of use for the HeAR model were followed. Every use of models and data from third parties complied with the relevant licensing requirements.

Statement of informed consent

The study used publicly available data and did not include any recruitment or gathering of individual participants. Therefore, informed consent was not necessary.

Funding

This research received no external funding

References

[1] Momtazmanesh S, Moghaddam SS, Ghamari SH, Rad EM, Rezaei N, Shobeiri P, et al. Global burden of chronic respiratory diseases and risk factors, 1990–2019: an update from the Global Burden of Disease Study 2019. EClinicalMedicine [Internet]. 2023 [cited 2026 May 19];59. Available from: https://www.thelancet.com/journals/eclinm/article/PIIS2589-5370(23)00113-X/fulltext

[2] Meghji J, Mortimer K, Agusti A, Allwood BW, Asher I, Bateman ED, et al. Improving lung health in low-income and middle-income countries: from challenges to solutions. The Lancet. 2021;397(10277):928–40.

[3] Kim Y, Hyon Y, Jung SS, Lee S, Yoo G, Chung C, et al. Respiratory sound classification for crackles, wheezes, and rhonchi in the clinical field using deep learning. Sci Rep. 2021;11(1):17186.

[4] Ehtesham A, Kumar S, Singh A, Khoei TT. Pediatric Asthma Detection with Google's HeAR Model: An AI-Driven Respiratory Sound Classifier. In: 2025 IEEE World AI IoT Congress (AIIoT) [Internet]. IEEE; 2025 [cited 2026 May 19]. p. 0103–9. Available from: https://ieeexplore.ieee.org/abstract/document/11105281/

[5] Garcia-Mendez JP, Lal A, Herasevich S, Tekin A, Pinevich Y, Lipatov K, et al. Machine learning for automated classification of abnormal lung sounds obtained from public databases: a systematic review. Bioengineering. 2023;10(10):1155.

[6] Zhang Y, Huang Q, Sun W, Chen F, Lin D, Chen F. Research on lung sound classification model based on dual-channel CNN-LSTM algorithm. Biomed Signal Process Control. 2024;94:106257.

[7] Wanasinghe T, Bandara S, Madusanka S, Meedeniya D, Bandara M, Díez IDLT. Lung sound classification with multi-feature integration utilizing lightweight CNN model. IEEE Access. 2024;12:21262–76.

[8] Huang DM, Huang J, Qiao K, Zhong NS, Lu HZ, Wang WJ. Deep learning-based lung sound analysis for intelligent stethoscope. Mil Med Res. 2023 Sep 26;10(1):44. doi:10.1186/s40779-023-00479-3

[9] Sreeram ASK, Ravishankar U, Sripada NR, Mamidgi B. Investigating the potential of MFCC features in classifying respiratory diseases. In: 2020 7th International Conference on Internet of Things: Systems, Management and Security (IOTSMS) [Internet]. IEEE; 2020 [cited 2026 May 19]. p. 1–7. Available from: https://ieeexplore.ieee.org/abstract/document/9340166/

[10] Alghamdi NS, Zakariah M, Karamti H. A deep CNN-based acoustic model for the identification of lung diseases utilizing extracted MFCC features from respiratory sounds. Multimed Tools Appl. 2024 Mar 12;83(35):82871–903. doi:10.1007/s11042-024-18703-0

[11] Biermann C, Han J, Mascolo C. Adapting Audio Foundation Models for Heart Sound Analysis [Internet]. 2025 [cited 2026 May 19]. Available from: https://cinc.org/archives/2025/pdf/CinC2025-253.pdf

[12] Mo A, Gui E, Fletcher RR. Use of voluntary cough sounds and deep learning for pulmonary disease screening in low-resource areas. In: 2022 IEEE Global Humanitarian Technology Conference (GHTC) [Internet]. IEEE; 2022 [cited 2026 May 19]. p. 242–9. Available from: https://ieeexplore.ieee.org/abstract/document/9911027/

[13] Dangi RR, Sharma A, Vageriya V. Transforming Healthcare in Low-Resource Settings With Artificial Intelligence: Recent Developments and Outcomes. Public Health Nurs. 2025 Mar;42(2):1017–30. doi:10.1111/phn.13500

[14] Tawfik M, Al-Zidi NM, Fathail I, Nimbhore S. Asthma Detection System: Machine and Deep Learning-Based Techniques. In: Pandit M, Gaur MK, Rana PS, Tiwari A, editors. Artificial Intelligence and Sustainable Computing [Internet]. Singapore: Springer Nature Singapore; 2022 [cited 2026 May 19]. p. 207–18. (Algorithms for Intelligent Systems). Available from: https://link.springer.com/10.1007/978-981-19-1653-3_16 doi:10.1007/978-981-19-1653-3_16

[15] Google. Hugging Face [Internet]. Higging Face; 2026 [cited 2026 May 19]. Google's Health Accoustics Representation Foundation Moddel (HeAR). Available from: https://huggingface.co/google/hear

[16] McFee B, Raffel C, Liang D, Ellis D, McVicar M, Battenberg E, et al. librosa: Audio and Music Signal Analysis in Python. In. Austin, Texas; 2015 [cited 2026 May 19]. p. 18–24. Available from: https://doi.curvenote.com/10.25080/Majora-7b98e3ed-003 doi:10.25080/Majora-7b98e3ed-003

[17] Bounab R, Zarour K, Guelib B, Khlifa N. Enhancing Medicare Fraud Detection Through Machine Learning: Addressing Class Imbalance With SMOTE-ENN. IEEE Access. 2024;12:54382–96. doi:10.1109/ACCESS.2024.3385781

[18] Scikit-Learn. scikit-learn: machine learning in Python — scikit-learn 1.8.0 documentation [Internet]. [cited 2026 May 19]. Available from: https://scikit-learn.org/stable/

[19] Borisov V, Leemann T, Seßler K, Haug J, Pawelczyk M, Kasneci G. Deep neural networks and tabular data: A survey. IEEE Trans Neural Netw Learn Syst. 2022;35(6):7499–519.

[20] Enda K, Oda Y, Tanei Z, Satoh K, Motegi H, Terasaka S, et al. Transfer Learning Strategies for Pathological Foundation Models: A Systematic Evaluation in Brain Tumor Classification. Pathol Int. 2026 Feb;76(2):e70098. doi:10.1111/pin.70098

[21] Ghrabli S, Elgendi M, Menon C. Identifying unique spectral fingerprints in cough sounds for diagnosing respiratory ailments. Sci Rep. 2024;14(1):593.

[22] Chen H, Yuan X, Pei Z, Li M, Li J. Triple-classification of respiratory sounds using optimized s-transform and deep residual networks. IEEE Access. 2019;7:32845–52.

[23] Naqvi SZH, Choudhry MA. An Automated System for Classification of Chronic Obstructive Pulmonary Disease and Pneumonia Patients Using Lung Sound Analysis. Sensors. 2020 Nov 14;20(22):6512. doi:10.3390/s20226512