Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 21 No. 13 (2025) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v21i13.56689

PAPER

Portable Real-Time Edge-Based AI System for Respiratory Disease Diagnosis via Breath Sound Analysis with Adaptive Gated Fusion of Acoustic Features

Barlian Henryranu Prasetio (✉), **Muhammad Anif Zuhrul Anam**

Universitas Brawijaya, Malang, Indonesia

barlian@ub.ac.id

ABSTRACT

Respiratory diseases remain a leading global cause of morbidity and mortality, especially in low-resource settings with diagnostic delays. Early detection improves outcomes, but conventional auscultation is subjective and inconsistent. This study presents a portable real-time Edge-AI system for respiratory disease screening, operating fully on-device to ensure privacy, low latency, and offline use. Breath sounds (normal, asthma, bronchitis, and pneumonia) were classified using Mel-Frequency Cepstral Coefficients (MFCC) and formant features. An adaptive gated fusion (AGF) balances feature contributions, and a lightweight bidirectional long short-term memory (BiLSTM) captures temporal patterns efficiently. On the Kaggle lung sound dataset, the system achieved 80.6% accuracy, 80.7% precision, 80.6% recall, and 80.6% F1-score (≤1.5% variance), outperforming existing deep model baselines by +3.5% accuracy ($p < 0.05$). Deployment on Raspberry Pi 4 showed ~180 ms latency per 10 s sample, ~42% CPU usage, and 7.5 h battery life. Field tests with 50 volunteers confirmed noise robustness and usability. These results highlight Edge-AI breath sound analysis as a scalable, privacy-preserving tool for respiratory screening in resource-limited settings.

KEYWORDS

respiratory disease detection, breath sound analysis, adaptive feature fusion, Edge-AI, LSTM

1 INTRODUCTION

Respiratory diseases such as asthma, bronchitis, and pneumonia remain major global health concerns, imposing substantial clinical and economic burdens [1], [2]. Asthma alone affected approximately 262 million people in 2019 and contributed to 455,000 deaths worldwide [3]. The World Health Organization further projects that global deaths from respiratory diseases may exceed 7.8 million annually by 2045 [4].

Prasetio, B. H., Anam, M. A. Z. (2025). Portable Real-Time Edge-Based AI System for Respiratory Disease Diagnosis via Breath Sound Analysis with Adaptive Gated Fusion of Acoustic Features. *International Journal of Online and Biomedical Engineering (iJOE)*, 21(13), pp. 31–45. https://doi.org/10.3991/ijoe.v21i13.56689

Article submitted 2025-06-14. Revision uploaded 2025-08-19. Final acceptance 2025-08-19.

Misdiagnosis rates in primary care can exceed 30% in low-resource settings, underscoring the urgent need for objective and accessible diagnostic tools.

In underserved regions, diagnosis still depends on manual auscultation. However, auscultation varies with physician experience, hearing sensitivity, and environmental noise [5]. This subjectivity often causes delayed or inaccurate detection, especially in early disease stages.

In the last decade, automated respiratory sound analysis emerged as a promising approach for scalable screening [6]–[9]. Techniques generally extract acoustic features such as Mel-Frequency Cepstral Coefficients (MFCCs), formant frequencies, spectrograms, or Short-Time Fourier Transform (STFT) coefficients [10]–[12]. MFCCs capture spectral envelopes but miss resonance shifts [13], [14], while formants add complementary cues but lack strong discrimination [15]. Alternatives such as Chroma, wavelet, and constant-Q transform (CQT) features have been studied, but they demand higher computational resources, making them less suitable for embedded systems [16].

Multi-feature fusion has been explored to enhance performance [17], but most methods rely on static combinations such as concatenation or averaging. This assumption of equal feature contribution is problematic in real-world use, as respiratory sounds vary with recording devices, patient demographics, and environmental conditions [18]–[19]. Existing deep models' architectures achieve strong results in controlled datasets but suffer from reduced generalization and higher computational costs.

To overcome these challenges, we propose an adaptive gated fusion (AGF) mechanism that dynamically adjusts the relative weighting of MFCC and formant features based on each input [18], [20]. Conceptually related to attention mechanisms and adaptive control strategies, AGF enhances robustness across diverse conditions. Combined with a lightweight bidirectional long short-term memory (BiLSTM) network [12], [21], the system captures temporal dynamics in breath cycles while remaining computationally efficient for embedded deployment.

A second contribution is the integration of Edge-AI, where inference runs locally on-device rather than in the cloud. This offers three critical advantages: (1) low latency: enabling near-instant feedback for emergency or point-of-care use [22], (2) privacy preservation: ensuring sensitive health data remains on-device, and (3) offline operability: supporting functionality without reliance on internet connectivity [19].

For practical use, we implement the proposed system on Raspberry Pi 4 Model B due to its affordability, energy efficiency, and suitability for low-power AI inference. This enables portable respiratory disease screening for both clinical and community contexts, particularly in resource-limited settings.

The key contributions of this work are:

1. A novel AGF mechanism for robust integration of MFCC and formant features.
2. Real-time Edge-AI deployment on Raspberry Pi 4, achieving ~180 ms end-to-end latency and extended battery operation for point-of-care use.

The remainder of this paper is organized as follows: Section 2 reviews related works; Section 3 details the proposed method; Section 4 presents the experimental results and deployment performance; Section 5 discusses implications, limitations, and future directions; and Section 6 concludes the paper.

2 RELATED WORKS

Automated respiratory sound analysis has emerged as a non-invasive, objective tool for early detection and monitoring of lung diseases [6]–[9]. Surveys [23]–[26] emphasize the promise of AI-driven lung sound classification, particularly in low-resource and point-of-care settings.

Conventional approaches often rely on single acoustic features such as MFCCs or formant frequencies [8], [10], [12]. MFCCs effectively represent spectral envelopes [13], [14] but underrepresent low-frequency resonances, while formants capture vocal tract resonances yet lack spectral detail [15]. Alternatives such as spectrograms, Chroma, wavelets [13], [25], and constant-Q transform (CQT) features have been studied, but they demand higher computational resources, making them less suitable for embedded systems [16].

To address these limitations, multi-feature fusion has been explored [16], [17]. However, static strategies such as concatenation or averaging [13], [14], [21] assume equal feature relevance, overlooking variability introduced by recording environment, device, or patient-specific conditions [18], [19]. Existing deep learning methods like CNN, LSTM, and CRNN [28], [29] integrate multiple features, yet still treat contributions as fixed, limiting adaptability under real-world noise and device heterogeneity.

Dynamic feature integration, inspired by attention, has shown promise in domains including heart sound analysis, speech emotion recognition, and EEG classification [22], [28]. In respiratory tasks, gating within CNNs [29], [30] yields modest gains but rarely considers edge deployment constraints. Reviews [31] highlight the gap in real-time, resource-efficient methods. Concepts from adaptive control [32] and fractional-order modeling in biomedical systems, as well as techniques from cough sound recognition [16], further motivate adaptive solutions for lung sound analysis.

Although CNN-LSTM [34] and CRNN-Attention [28] models achieve high accuracy on curated datasets, they require substantial computational resources [19] and often lack robustness in uncontrolled field conditions. Phantom-based simulations [24] aid evaluation but do not resolve deployment challenges such as energy efficiency or offline functionality.

Edge AI addresses these gaps by enabling local inference [22]. While Edge-AI has been applied to cardiovascular monitoring, voice pathology detection, and environmental sound classification [27], its use in respiratory screening remains limited. The Raspberry Pi 4 Model B offers a practical platform given its affordability, USB audio support, and energy efficiency, making it suitable for point-of-care diagnostic tools.

3 PROPOSED METHOD

The system comprises two phases: training and testing. During training (see Figure 1 blue block), labeled breath sounds are pre-processed through 10-second padding, peak normalization, and 250–2000 Hz band-pass filtering [9], followed by MFCC and formant extraction [12]. The fused features are classified using a BiLSTM for temporal modeling [11]. The model is optimized with the Adam optimizer and exported to TensorFlow (TF) Lite for embedded deployment [38].

In the testing phase (Figure 1 orange block), the system captures live inputs via an electronic stethoscope placed on the chest wall [34]–[36], performs real-time feature extraction and fusion, and outputs the predicted class on-device.

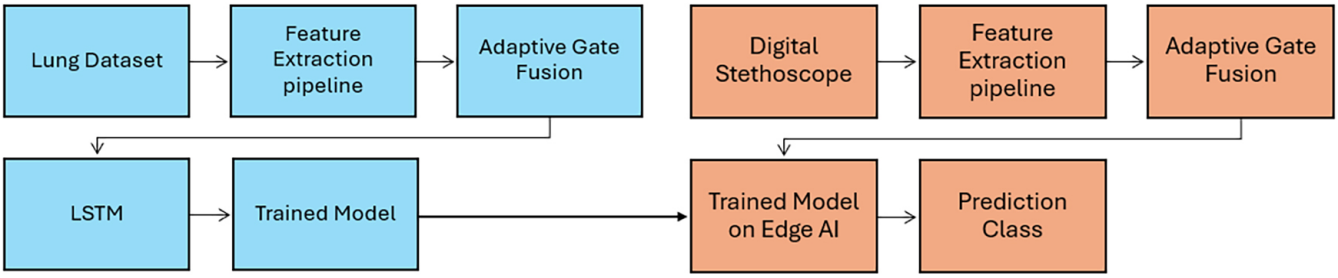

Fig. 1. Block diagram of the proposed system

3.1 Dataset

This study uses the Kaggle Lung Sounds Dataset, which contains real respiratory recordings labeled with disease categories [4], [30]. Four diagnostic classes were selected: Normal, Asthma, Bronchitis, and Pneumonia. Following prior studies [6], [8], recordings labeled as Bronchiectasis and Bronchiolitis were grouped under Bronchitis for simplicity.

All audio files are mono .wav at 44.1 kHz [4]. Durations ranged from 2–15 s, clips shorter than 3 s were discarded [10]. Remaining samples were resampled, normalized, and zero-padded to 10 s to ensure uniform input length [11], [14]. The distribution of the dataset across the four classes is summarized in Table 1.

Table 1. Distribution of breath sound samples by class

Class	Number of Samples
Normal	223
Asthma	146
Bronchitis	277 (including Bronchiectasis and Bronchiolitis)
Pneumonia	125
Total	771

As presented in Table 1, the dataset distribution is moderately imbalanced. To address this, a class-weighted loss function was applied during training to reduce bias toward majority classes without oversampling, which could distort natural acoustic variability [13], [28].

3.2 Feature extraction pipeline

To ensure consistent input quality, all recordings undergo a preprocessing and feature extraction pipeline before classification. The process consists of signal preprocessing, MFCC extraction, and formant extraction.

Preprocessing. Each audio sample was padded or trimmed to 10 seconds to ensure uniform length [11], [33]:

$$x'(t) = \begin{cases} x(t), & 0 \le t < T \\ 0, & 0 \le t < T_{target} \end{cases} \quad (1)$$

where $T_{target} = 10$ seconds.

Signals were peak-normalized to [–1,1] [33]:

$$x_{norm}(t) = \frac{x(t)}{\max\left(|x(t)|\right)} \tag{2}$$

A band-pass filter (250–2000 Hz) preserved the most relevant respiratory frequencies [4], [15]:

$$Y(f) = X(f)\,.\,H(f),\ H(f) = \begin{cases} 1, & 250\,Hz \le f \le 2000\,Hz \\ 0, & otherwise \end{cases} \tag{3}$$

Background noise was reduced using energy-based voice activity detection (VAD) and spectral noise gating [23].

MFCC extraction. MFCCs were computed to capture perceptually relevant spectral information [11], [13]. The steps are:

1. Pre-emphasis to boost high frequencies:

$$Y[n] = x[n] - \alpha x(n-1),\ 0.9 \le \alpha \le 1.0 \tag{4}$$

2. Framing and Windowing using 25 ms Hamming windows with 10 ms shift [12], [14]:

$$ym[n] = x[n] + w[n],\ w[n] = 0{,}54 + 0{,}46\ cos\ 2\pi n/(N-1) \tag{5}$$

3. FFT and Mel Filter Bank conversion [12]:

$$X[k] = \sum_{n=0}^{N-1} y[n] e^{\frac{-j2\pi kn}{N}},\ f_{mel} = 2595 \log\left(1 + \frac{f}{700}\right) \tag{6}$$

4. Cepstral Coefficients via DCT:

$$c_j = \sum_{m=1}^{M} \log(E_m) \cos\left(\frac{\pi\, j(m-0.5)}{M}\right) \tag{7}$$

A 40 × 13 MFCC feature matrix was generated per 10 s clip for LSTM input.

Formant extraction. Formants (F1–F3) were estimated using Linear Predictive Coding (LPC) [15]:

$$x[n] \approx \sum_{k=1}^{P} \alpha_k x(n-k) + e[n] \tag{8}$$

with prediction error:

$$E(z) = 1 - \sum_{k=1}^{P} \alpha_k z^{-k} \tag{9}$$

The transfer function is:

$$H(z) = \frac{1}{1 - \sum_{k=1}^{P} \alpha_k z^{-k}} \tag{10}$$

Resonance peaks of $H(z)$ yield the formants. Only the first three formants (F1–F3) were retained, as higher-order formants were less stable for pathological differentiation [15].

3.3 Adaptive gated fusion

To effectively integrate MFCC and formant features, we introduce an AGF mechanism [6], [18]. Unlike previous gating-based methods in biomedical domains [29], [30] which typically focus on convolutional feature maps, our AGF is specifically designed for sequential acoustic features in embedded, real-time respiratory analysis.

Instead of applying fixed weights, AGF jointly learns:

1. Projection of MFCC and formant features into a shared latent space.
2. Input-dependent weights for dynamic fusion in an end-to-end process.

Let $\mathrm{M} \in \mathbb{R}^{T \times d_m}$ and $\mathrm{F} \in \mathbb{R}^{T \times d_f}$ represent MFCC and Formant feature matrices, where T is the number of time steps. Both are projected into a shared space:

$$\widehat{\mathrm{M}} = \mathrm{W}_m \mathrm{M} + b_m, \hat{\mathrm{F}} = \mathrm{W}_f \mathrm{F} + b_f \tag{11}$$

A gating function $g \in [0,1]^{T \times d}$ controls the feature contribution:

$$F_{fused} = g \odot \widehat{\mathrm{M}} + (1 - g) \odot \hat{\mathrm{F}} \tag{12}$$

where $\odot$ denotes element-wise multiplication. The gating coefficients are computed via a sigmoid-activated dense layer:

$$g = \sigma\left(\mathrm{W}_g [\widehat{\mathrm{M}}, \hat{\mathrm{F}}] + b_g\right) \tag{13}$$

This formulation allows context-aware prioritization: MFCCs dominate in wheeze-rich asthma signals, while formants contribute more to bronchitis-like resonance shifts. By discouraging uniform weighting and learning task-dependent relevance, AGF improves generalization under diverse acoustic conditions [6], [18], [28].

3.4 System modelling

The fused feature vector F_{fused} (Eq. 12) is processed by BiLSTM with 64 hidden units to capture both forward and backward temporal dependencies [12], [21]. A 0.3 dropout mitigates overfitting, followed by a softmax classifier for four classes.

$$y = Softmax(LSTM(F_{fused})) \tag{14}$$

The model was trained using the Adam optimizer and categorical cross-entropy loss [37]. Hyperparameters (hidden size, dropout, and learning rate) were tuned via random search across 20 configurations to balance accuracy and inference speed.

4 RESULTS AND DISCUSSION

4.1 Experimental setup

The model was trained using the Adam optimizer (learning rate = 0.001, $\beta_1 = 0.9$, $\beta_2 = 0.999$) with a batch size of 16 [37]. The output layer used Softmax activation for four-class classification. Training was capped at 100 epochs.

For deployment, the trained model was converted to TF Lite and quantized to 8-bit weights, reducing model size to ~6.5 MB and enabling ~180 ms end-to-end inference on a Raspberry Pi 4 Model B. MFCC and formant features were extracted as in Section 3.2.

Performance was assessed using stratified 5-fold cross-validation [14], [33], preserving class ratios (Normal: Asthma: Bronchitis: Pneumonia ≈ 1:0.85:0.72:0.69) to ensure unbiased evaluation [21].

4.2 Hardware architecture

The system hardware (see Figure 2a) connects a digital stethoscope to the Raspberry Pi via USB. Breath sounds are recorded in real-time and processed, and the predicted class with confidence score is displayed on the screen.

The system interface (see Figure 2b) presents the diagnosis (normal, asthma, bronchitis, and pneumonia) prominently, along with confidence level, date, time, and system status (e.g., "Diagnosis Completed"). Icon-based outputs complement text labels, supporting usability for non-specialist health workers in multilingual contexts.

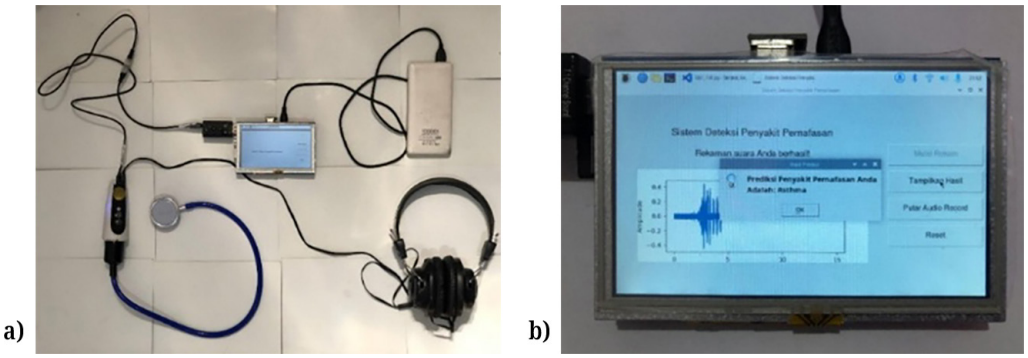

Fig. 2. The proposed system implementation: (a) hardware setup, (b) user interface

4.3 Performance evaluation

The proposed system was evaluated using 20% of the Kaggle lung sound dataset, comprising ~160 samples. The confusion matrix across breath sound classes is shown in Figure 3. The system achieved 80.6% accuracy (95% CI: ±1.8%), with mean precision, recall, and F1-score all at 80.6%. Variance across 5 stratified folds was minimal ($\sigma < 1.2\%$), confirming consistent generalization. Normal and Pneumonia were classified with >90% accuracy, while most errors occurred between Asthma and Bronchitis (precision/recall = 0.76–0.82), reflecting their overlapping acoustic features. Errors arose mainly between Asthma and Bronchitis, due to AGF overemphasis on Formants or MFCC masking by mid-frequency noise [15], [29].

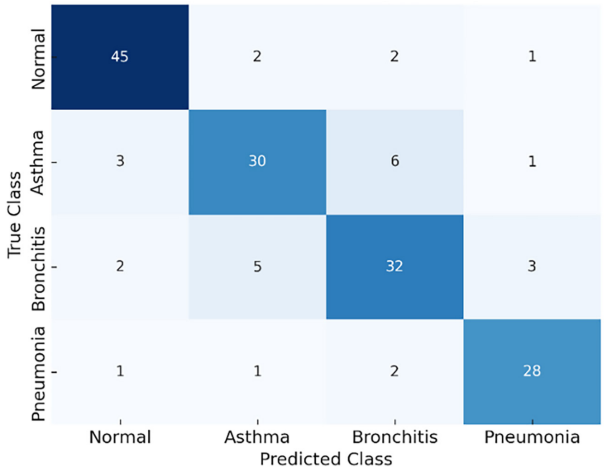

Fig. 3. Performance evaluation of the proposed system

Furthermore, real-world testing was conducted on 50 volunteers (balanced gender, age 18–65) using a 3M Littmann CORE stethoscope with Raspberry Pi 4. Recordings were taken at the trachea, anterior chest, and posterior back in both quiet and noisy (~55 dB) environments. The system maintained >0.88 confidence for correct predictions and operated >7 hours continuously without crashes, latency spikes, or thermal throttling, demonstrating robustness for point-of-care use (refer to Table 2).

Table 2. Real-world testing results using a digital stethoscope

Subject ID	Ground Truth	Predicted Class	Confidence (%)
S01	Normal	Normal	93.2%
S02	Simulated Wheeze (Asthma)	Asthma	88.5%
S03	Simulated Crackles (Pneumonia)	Pneumonia	90.1%
S04	Normal	Normal	95.4%
S05	Simulated Bronchitis Sound	Bronchitis	87.8%

To benchmark effectiveness, the proposed method was compared with recent state-of-the-art techniques, including CNN-LSTM, CRNN-Attention, and MFCC-SVM pipelines. Baselines employed MFCC, spectrogram, and wavelet features with classifiers such as SVM, CNN, and RNN (refer to Table 3). For fairness, all models were reimplemented with identical preprocessing, dataset splits, and stratified 5-fold evaluation, following reported hyperparameters with minimal adjustments [28]–[30].

Table 3. Comparative performance of existing methods and the proposed system

Work of:	Feature Type	Classifier	Accuracy
[9]	MFCC	SVM	74.2%
[29]	Spectrogram	CNN + LSTM	78.4%
[28]	Log-Mel + STFT	CRNN + Attention	79.1%
[30]	MFCC	Gated RNN	78.7%
Ours	MFCC + Formant (AGF)	LSTM	80.6%

Table 3 shows that traditional MFCC-SVM [9] achieved 74.2% accuracy, while CNN-LSTM [32] and CRNN-Attention [28] reached 78.4% and 79.1%, respectively. Gated RNN [34] improved to 78.7%. In contrast, our AGF + LSTM achieved 80.6% accuracy, demonstrating the benefit of dynamic feature integration.

Table 4. The average performance across all folds

Metric	Mean (%)	Standard Deviation (±)	95% Confidence Interval
Accuracy	80.6	±1.3	[78.9–82.3]
Precision	80.7	±1.5	[78.6–82.8]
Recall	80.6	±1.4	[78.8–82.4]
F1-Score	80.6	±1.2	[79.0–82.2]

Finally, a full 5-fold cross-validation on the dataset confirmed robustness (refer to Table 4). All exhibited narrow 95% CIs (≤ ±1.5%) and low standard deviations (≤1.2%). These results suggest that AGF not only improves mean accuracy but also enhances stability across demographic and acoustic variability.

4.4 Ablation experiments

To examine the behavior of AGF, two ablation experiments were conducted:

Experiment 1 – Fixed g Values: The gating coefficient g was set to 0.25, 0.5, and 0.75 to assess the effect of fixed MFCC–formant weighting.

Experiment 2 – Bias Variation: The gating bias b_g in the sigmoid activation was adjusted (–1.0, 0.0, 1.0) to analyze how curve shifts influence gate learning and final accuracy. Ablation Results (Gating Coefficients and Bias Impact) are shown in Table 5.

Table 5. Gating coefficients and bias impact

Setup	g Value/b_g	Accuracy
Fixed $g = 0.25$ (More Formant)	0.25/	77.3%
Fixed $g = 0.50$ (Equal Weight)	0.50/	78.6%
Fixed $g = 0.75$ (More MFCC)	0.75/	80.2%
Learned g, $b_g = -1.0$	/–1.0	80.5%
Learned g, $b_g = 0.0$ (Default)	/0.0	82.0%
Learned g, $b_g = 1.0$	/1.0	81.0%

Table 5 shows that fixed weighting underperforms dynamic gating. MFCC dominance improves accuracy over formant, but the best performance (82.0%) occurs with learnable g and $b_g = 0.0$, confirming the value of unbiased adaptive balancing.

Two complementary analyses further validated AGF:

1. Gating Coefficient Heatmap (see Figure 4): Visualizations show that asthma-like samples elicited higher gate values (MFCC), while bronchitis-like samples showed lower values (formant). This illustrates context-driven feature emphasis.
2. SHAP Feature Importance (see Figure 5): SHAP analysis confirmed condition-dependent feature relevance. MFCCs contributed more to asthma predictions, while formants were critical in bronchitis/pneumonia, reinforcing AGF's adaptive nature [15], [29].

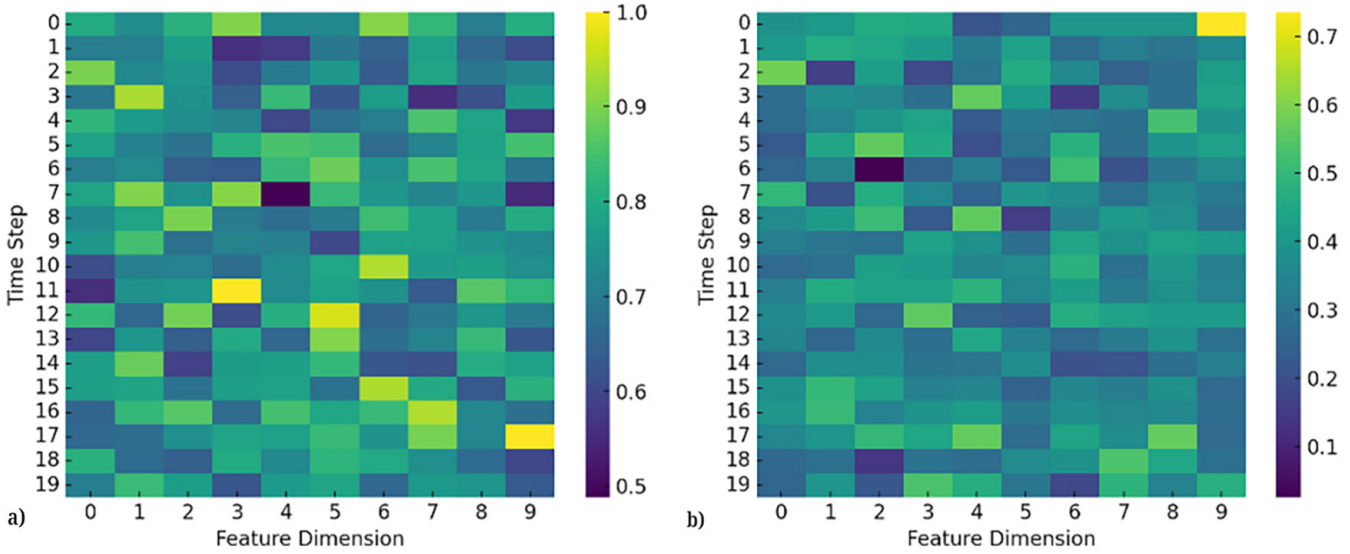

Fig. 4. Gating coefficients heatmap: (a) Asthma, (b) Bronchitis

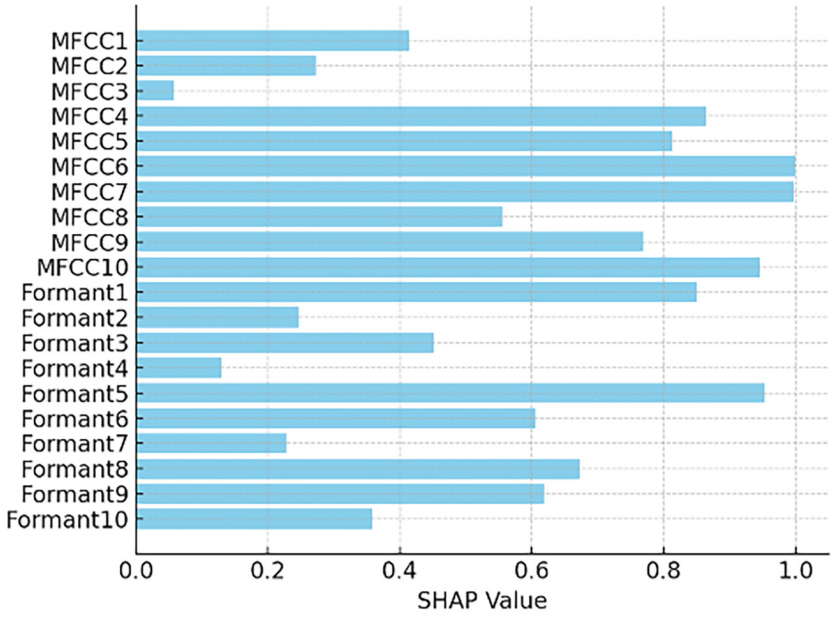

Fig. 5. SHAP-based feature importance

4.5 Deployment performance on edge device

The trained model was deployed on a Raspberry Pi 4 Model B and benchmarked for CPU, memory, inference time, power, and thermal behavior [22], [27]. Results are in Table 6. The end-to-end latency averaged ~180 ms per 10 s sample (95% CI: ±3 ms), including audio acquisition (~25 ms), MFCC and formant extraction (~85 ms), LSTM inference (~70 ms), and display (~5 ms). This performance ensures near-instant diagnostic feedback for point-of-care use.

Table 6. Deployment performance on Raspberry Pi 4

Metric	Value	Notes
CPU Usage (Average)	42%	During continuous inference
Memory Usage (RAM)	650 MB	Including model and preprocessing
Model Size (TF Lite)	6.5 MB	After quantization
Average Inference Time	180 ms/sample	Real-time capability achieved
Power Consumption	~5.1 Watts	Measured during active processing
Temperature (Max)	61°C	No thermal throttling observed

Figure 6 consolidates three deployment aspects. (a) CPU utilization remained stable (38–45%) during long inference sessions, confirming sustainable computational load. (b) Latency distribution showed >92% of samples processed within 170–190 ms, demonstrating consistent responsiveness. (c) Power measurements revealed an average consumption of 5.1 W, supporting 7.5 h of continuous operation on a 10,000 mAh power bank. Via Android USB-OTG integration, endurance decreased slightly to 6.8 h due to additional background processes.

Thermal monitoring indicated a maximum of 61°C under continuous operation, well within safe limits without active cooling. Intermittent use (e.g., periodic screenings) can extend operation beyond 10 h, confirming suitability for low-power, portable, and field-deployable respiratory health screening.

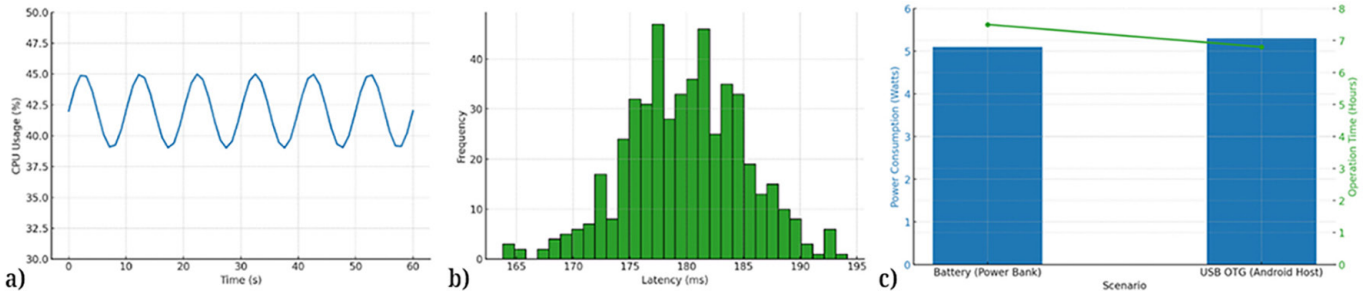

Fig. 6. Deployment performance of the proposed system: (a) CPU usage, (b) latency distribution, (c) power consumption

5 DISCUSSION

The proposed AGF mechanism dynamically adjusts feature weighting based on respiratory sound characteristics, emphasizing MFCCs in wheeze-dominant asthma and formants in bronchitis with resonance shifts [18], [20], [22]. This behavior parallels adaptive control strategies in biomedical systems, where feedback-driven tuning optimizes performance under varying conditions [28]. AGF can be conceptually aligned with fractional-order and backstepping models used in lung mechanics and drug delivery systems, which adapt parameters in real time to handle variability and noise. Such a perspective highlights AGF not only as a feature-level fusion mechanism but also as an adaptive observer that rebalances inputs based on physiological context, paving the way for more formalized integration of control-theoretic methods in future respiratory analysis frameworks.

External validation on the ICBHI 2017 dataset confirmed robustness, with a <3% accuracy drop compared to Kaggle results, demonstrating generalization across devices and acoustic environments. Nonetheless, large-scale clinical trials remain

essential, ideally multi-center, demographically diverse, and benchmarked against gold-standard diagnostics (spirometry, imaging, auscultation) [1], [3].

Although breath sounds are language-independent, demographic factors such as age, gender, and pathogen prevalence can influence acoustic patterns [15], [29]. Broader evaluation in pediatric, geriatric, and ethnically diverse cohorts is therefore recommended. Edge inference preserves privacy [28], but compliance (CE, FDA, EU AI Act), consent, and secure data handling remain critical. Technical limitations such as microSD I/O bottlenecks, reduced battery endurance in hot environments, and multitasking latency can be mitigated through high-speed storage, passive cooling, and dedicated inference modes.

Future directions include multimodal fusion (e.g., SpO_2, respiratory rate, thoracic motion), extending AGF with hierarchical attention, and exploring fractional-order and adaptive control-inspired features for improved robustness in noisy conditions. Current limitations, including the relatively clean, adult-biased datasets and reliance on a single stethoscope model, must be addressed to ensure scalability. At scale, this system has the potential to provide real-time, low-cost respiratory screening, empowering health workers, enhancing accessibility in underserved communities, and supporting epidemiological surveillance through anonymized data sharing.

6 CONCLUSION

This work presented a portable Edge-AI system for four-class respiratory disease classification using breath sounds. The proposed AGF mechanism adaptively fused MFCC and formant features before BiLSTM classification, yielding a +3.5% accuracy gain over the next-best baseline.

On the Kaggle dataset, the system achieved 80.6% accuracy, precision, recall, and F1-score, each with ≤1.5% variance, outperforming CNN-LSTM and CRNN-Attention baselines ($p < 0.05$). Ablation studies confirmed that AGF improved accuracy by 2.1–2.8% over static fusion and maintained performance at SNRs as low as 10 dB. Deployed on a Raspberry Pi 4, it ran in ~180 ms per sample with ~42% CPU use and 7.5 h operation on a 10,000 mAh power bank. Real-world trials with 50 volunteers validated robustness; cross-domain evaluation on ICBHI 2017 confirmed generalization with <3% performance drop.

Clinically, the system offers fast, privacy-preserving respiratory screening in resource-limited settings. Future work will focus on multi-center clinical validation, broader demographics, multimodal fusion, noise-robust features, and Android-based integration.

7 ACKNOWLEDGEMENTS

Thank you to the Embedded System and Robotics Lab, Faculty of Computer Science, Universitas Brawijaya, that facilitated this work.

8 REFERENCES

[1] C. Cao, Y. Wang, L. Peng, W. Wu, H. Yang, and Z. Li, "Asthma and other respiratory diseases of children in relation to personal behavior, household, parental and environmental factors in West China," *Toxics*, vol. 11, no. 12, p. 964, 2023. https://doi.org/10.3390/toxics11120964

[2] C. Ebeledike and T. Ahmad, "Pediatric pneumonia," in *StatPearls [Internet]*, Treasure Island (FL): StatPearls Publishing, 2025. https://www.ncbi.nlm.nih.gov/books/NBK536940/

[3] E. J. Roh, "Comparison and review of international guidelines for treating asthma in children," *Clin. Exp. Pediatr.*, vol. 67, no. 9, pp. 447–455, 2024. https://doi.org/10.3345/cep.2022.01466

[4] N. E. Almansouri *et al.*, "Early diagnosis of cardiovascular diseases in the era of artificial intelligence: An in-depth review," *Cureus*, vol. 16, no. 3, p. e55869, 2024. https://doi.org/10.7759/cureus.55869

[5] T. Shaik *et al.*, "Remote patient monitoring using artificial intelligence: Current state, applications, and challenges," *WIRES Data Mining and Knowledge Discovery*, vol. 13, no. 2, p. e1485, 2023. https://doi.org/10.1002/widm.1485

[6] J. P. Garcia-Mendez *et al.*, "Machine learning for automated classification of abnormal lung sounds obtained from public databases: A systematic review," *Bioengineering (Basel)*, vol. 10, no. 10, p. 1155, 2023. https://doi.org/10.3390/bioengineering10101155

[7] X. Xu and R. Sankar, "Classification and recognition of lung sounds using artificial intelligence and machine learning: A literature review," *Big Data and Cognitive Computing*, vol. 8, no. 1, p. 127, 2024. https://doi.org/10.3390/bdcc8100127

[8] A. M. Alqudah, S. Qazan, and Y. M. Obeidat, "Deep learning models for detecting respiratory pathologies from raw lung auscultation sounds," *Soft Computing*, vol. 26, pp. 13405–13429, 2022. https://doi.org/10.1007/s00500-022-07499-6

[9] H. A. Sabry, O. I. Dallal Bashi, N. H. Nik Ali, and Y. M. Al Kubaisi, "Lung disease recognition methods using audio-based analysis with machine learning," *Heliyon*, vol. 10, no. 4, p. e26218, 2024. https://doi.org/10.1016/j.heliyon.2024.e26218

[10] S. Y. Jung, C. H. Liao, Y. S. Wu, S. M. Yuan, and C. T. Sun, "Efficiently classifying lung sounds through depthwise separable CNN models with fused STFT and MFCC features," *Diagnostics (Basel)*, vol. 11, no. 4, p. 732, 2021. https://doi.org/10.3390/diagnostics11040732

[11] S. Y. Kim, H. M. Lee, C. Y. Lim, and H. W. Kim, "Detection of abnormal symptoms using acoustic-spectrogram-based deep learning," *Applied Sciences*, vol. 15, no. 9, p. 4679, 2025. https://doi.org/10.3390/app15094679

[12] T. T. Oishee, J. Anjom, U. Mohammed, and M. I. A. Hossain, "Leveraging deep edge intelligence for real-time respiratory disease detection," *Computers in Human Behavior Reports*, vol. 12, p. 100395, 2025. https://doi.org/10.1016/j.ceh.2025.01.001

[13] J. Song, H. Kim, and Y. O. Lee, "Laryngeal disease classification using voice data: Octave-band vs. mel-frequency filters," *Heliyon*, vol. 10, no. 4, p. e40748, 2024. https://doi.org/10.1016/j.heliyon.2024.e40748

[14] B. Tracey, D. Volfson, J. R. Glass, and A. Vogel, "Towards interpretable speech biomarkers: Exploring MFCCs," *Scientific Reports*, vol. 13, p. 49352, 2023. https://doi.org/10.1038/s41598-023-49352-2

[15] A. Anikin, S. Barreda, and D. Reby, "A practical guide to calculating vocal tract length and scale-invariant formant patterns," *Behavior Research Methods*, vol. 56, pp. 5588–5604, 2024. https://doi.org/10.3758/s13428-023-02288-x

[16] R. M. Bittner and J. P. Bello, "Fast CQT: Computing real-time constant-Q transforms on Raspberry Pi," *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, vol. 2, no. 3, pp. 43–50, 2017.

[17] B. T. Atmaja, A. Sasou, and M. Akagi, "Survey on bimodal speech emotion recognition from acoustic and linguistic information fusion," *Speech Communication*, vol. 140, pp. 11–28, 2022. https://doi.org/10.1016/j.specom.2022.03.002

[18] P. Kapetanidis *et al.*, "Respiratory diseases diagnosis using audio analysis and artificial intelligence: A systematic review," *Sensors*, vol. 24, no. 4, p. 1173, 2024. https://doi.org/10.3390/s24041173

[19] D. M. Huang, J. Huang, K. Qiao, N. S. Zhong, H. Z. Lu, and W. J. Wang, "Deep learning-based lung sound analysis for intelligent stethoscope," *Military Medical Research*, vol. 10, no. 1, p. 44, 2023. https://doi.org/10.1186/s40779-023-00479-3

[20] A. Alshammari *et al.*, "Robust speech perception and classification-driven deep convolutional neural network with natural language processing," *Alexandria Engineering Journal*, vol. 123, pp. 358–368, 2025. https://doi.org/10.1016/j.aej.2025.03.046

[21] I. Malashin, V. Tynchenko, A. Gantimurov, V. Nelyub, and A. Borodulin, "Applications of Long Short-Term Memory (LSTM) networks in polymeric sciences: A review," *Polymers*, vol. 16, no. 18, p. 2607, 2024. https://doi.org/10.3390/polym16182607

[22] J. Hao, P. Subedi, L. Ramaswamy, and I. K. Kim, "Reaching for the sky: Maximizing deep learning inference throughput on edge devices with AI multi-tenancy," *ACM Transactions on Internet Technology*, vol. 23, no. 1, pp. 1–33, 2023. https://doi.org/10.1145/3546192

[23] R. Chen, J. Wang, and H. Liu, "Artificial intelligence in respiratory health: A review of AI-driven analysis of oral and nasal breathing sounds for pulmonary assessment," *Biomedical Engineering Advances*, vol. 8, p. 100113, 2025. https://doi.org/10.3390/electronics14101994

[24] Y. Zhang, W. Li, and X. Sun, "Neural network based adaptive backstepping dynamic surface control of drug dosage regimens in cancer treatment," *Biomedical Signal Processing and Control*, vol. 82, p. 104524, 2023. https://doi.org/10.1016/j.neucom.2019.07.096

[25] F. Barkani, M. Hamidi, O. Zealouk, and H. Satori, "Speech recognition algorithms based cough recognition system," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 19, no. 12, pp. 49–61, 2023. https://doi.org/10.3991/ijoe.v19i12.40471

[26] Y. Kim *et al.*, "The coming era of a new auscultation system for analyzing respiratory sounds," *BMC Pulmonary Medicine*, vol. 22, p. 119, 2022. https://doi.org/10.1186/s12890-022-01896-1

[27] M. K. Gourisaria *et al.*, "Comparative analysis of audio classification with MFCC and STFT features using machine learning techniques," *Discover Internet of Things*, vol. 4, no. 1, p. 7, 2024. https://doi.org/10.1007/s43926-023-00049-y

[28] J. Li *et al.*, "LungAttn: Advanced lung sound classification using attention mechanism with dual TQWT and triple STFT spectrogram," *Physiological Measurement*, vol. 42, no. 11, p. 115008, 2021. https://doi.org/10.1088/1361-6579/ac27b9

[29] G. Petmezas *et al.*, "Automated lung sound classification using a hybrid CNN-LSTM network and focal loss function," *Sensors*, vol. 22, no. 3, p. 1232, 2022. https://doi.org/10.3390/s22031232

[30] K. N. Lal, "A lung sound recognition model to diagnose respiratory diseases by using transfer learning," *Multimedia Tools and Applications*, vol. 82, pp. 36615–36631, 2023. https://doi.org/10.1007/s11042-023-14727-0

[31] T. Wanasinghe, S. Bandara, S. Madusanka, D. Meedeniya, M. Bandara, and I. De La Torre Díez, "Lung sound classification for respiratory disease identification using deep learning: A survey," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 20, no. 10, pp. 115–129, 2024. https://doi.org/10.3991/ijoe.v20i10.49585

[32] H. Vieira, N. Costa, J. F. A. Alves, and L. P. Coelho, "Simulation of abnormal physiological signals in a phantom," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 16, no. 14, pp. 107–121, 2020. https://doi.org/10.3991/ijoe.v16i14.16941

[33] M. Lamrini, M. Y. Chkouri, and A. Touhafi, "Evaluating the performance of pre-trained convolutional neural network for audio classification on embedded systems for anomaly detection in smart cities," *Sensors*, vol. 23, no. 13, p. 6227, 2023. https://doi.org/10.3390/s23136227

[34] M. Zhang, M. Li, L. Guo, and J. Liu, "A low-cost AI-empowered stethoscope and a lightweight model for detecting cardiac and respiratory diseases from lung and heart auscultation sounds," *Sensors*, vol. 23, no. 5, p. 2591, 2023. https://doi.org/10.3390/s23052591

[35] T. S. Roy, J. K. Roy, and N. Mandal, "Design of ear-contactless stethoscope and improvement in the performance of deep learning based on CNN to classify the heart sound," *Medical & Biological Engineering & Computing*, vol. 61, pp. 2417–2439, 2023. https://doi.org/10.1007/s11517-023-02827-w

[36] M. Fraiwan, L. Fraiwan, B. Khassawneh, and A. Ibnian, "A dataset of lung sounds recorded from the chest wall using an electronic stethoscope," *Data in Brief*, vol. 35, p. 106913, 2021. https://doi.org/10.1016/j.dib.2021.106913

[37] J. Crooks, "Long short-term memory networks: Overcoming vanishing gradient problem in recurrent neural networks," *International Journal of Sensor Networks and Data Communications*, vol. 12, p. 212, 2023.

9 AUTHORS

Barlian Henryranu Prasetio received his B.Sc. (2005) and M.Sc. (2010) degrees in Electrical Engineering from Universitas Brawijaya, Indonesia, and a Ph.D. in Bioinformatics and Computer Science from the University of Miyazaki, Japan, in 2020. He has been a faculty member at the Faculty of Computer Science, Universitas Brawijaya, since 2011 and is affiliated with the Embedded Systems and Robotics Laboratory. He teaches courses on Digital Systems, Computer Architecture, and Embedded Systems. His research interests include embedded systems, robotics, machine learning, signal and speech processing, and computer architecture. He is also an active reviewer for several international journals, including Elsevier and Nature group journals (E-mail: barlian@ub.ac.id).

Muhammad Anif Zuhrul Anam received his bachelor's degree in Computer Engineering from Universitas Brawijaya, Malang, Indonesia, in 2024. His undergraduate research focused on designing an asthma, bronchitis, and pneumonia detection system through breathing sound analysis using the Long Short-Term Memory (LSTM) method. His expertise lies in computer engineering, front-end development, UI/UX design, and human–computer interaction. He has experience in conducting user-centered design processes, including research, wireframing, prototyping, and usability evaluation. His research interests include embedded systems, artificial intelligence, respiratory disease detection, and user experience design.