Multi-Stage Respiratory Sound Analysis: Confidence-Driven Wheeze & Crackle Detection

Annapurna Kala, Mounya Elhilali

***Abstract*— *Objective*: Accurate detection of adventitious respiratory sounds, such as wheezes and crackles, is essential for diagnosing and managing respiratory conditions. This study introduces a multi-stage, confidence-driven framework for automated pediatric auscultation analysis, performing a three-way classification of normal, wheeze, and crackle sounds to improve diagnostic accuracy.**

***Methods*: We develop a comprehensive pipeline integrating anomaly-specific segment selection, segment-level classification, and confidence-based fusion. Our contrastive variational recurrent neural network (CVRNN) enhances feature extraction, while a confidence-weighted aggregation strategy refines final predictions. The system is validated using a diverse pediatric dataset from 742 subjects (aged 1-59 months) from seven countries.**

***Results*: The multi-level framework is evaluated across three stages. The anomaly-specific segment selection achieves 98.47 % recall, identifying adventitious regions. Next, segment-level classifiers improve sensitivity, achieving balanced accuracies of 72.15 % (wheeze) and 68.1 % (crackle). This performance surpasses state of the art systems on the same dataset and demonstrates enhanced balanced performance in detecting both crackle and wheeze sounds, which present different challenges to automated systems given their markedly different acoustic profiles. Finally, the confident-driven fusion outperforms traditional aggregation methods, yielding a final three-way classification of 62.12 %.**

***Conclusion*: Our confidence-based multi-stage approach enhances automated respiratory sound classification by prioritizing high-certainty segment predictions, aligning with expert physician annotations.**

***Significance*: This framework advances computer-aided respiratory diagnostics, improving early detection and monitoring of pediatric respiratory conditions. By integrating expert-inspired segmentation with machine learning-driven confidence estimation, it has potential to enhance clinical workflows and screening for pulmonary diseases, particularly in resource-limited settings.**

***Index Terms*— Auscultation Anomaly Detection, Confidence-Driven Decision making, Multi-stage lung audio event detection**

This work was supported in part by projects NIH R01HL163439 and ONR N00014-23-1-2086. The authors would like to thank the PERCH study group for guidance throughout the completion of this work, and to the patients and families enrolled in this study. We also thank Dr. Eric McCollum for scientific discussions and insights on interpreting auscultation signals and clinical data.

A. Kala and M. Elhilali are with the Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD 21218, USA. Correspondence: mounya@jhu.edu

I. Introduction

Pneumonia continues to be a significant global health challenge that accounts for around 15% of all deaths among children under the age of five, particularly in resource-limited settings [1], [2]. Early diagnosis plays a crucial role in improving patient outcomes, though there is no one symptom, exam or radiographic finding that is ever sufficient for diagnosing complex phenomena such as pneumonia. Still, recognizing practical considerations including usability, cost and wide deployment, respiratory auscultations serve as a crucial first line defense to assess lung function, with findings underscoring the clinical value of auscultations in pneumonia diagnosis [3]. In settings where advanced diagnostic tools are often scarce or unavailable [4], auscultation becomes even more indispensable, since sound-based assessment is inexpensive, quick, readily accessible, and noninvasive. Building on this foundation, computerized auscultation analysis can address gaps in clinical expertise across settings. By leveraging recent advances in machine learning and audio signal processing, computerized auscultation analysis can provide standardized, reproducible interpretations of lung sounds and ensure timely identification of patients who require additional medical intervention.

Typically, computerized auscultation analysis is framed as a detection task aimed at identifying anomalous breathing patterns for further clinical assessment [5]. In practice, these methods often focus on classifying wheezes and crackles, given their significance as primary adventitious sounds [6]–[9]. This approach aligns with the need for consistent training and inference methodologies in machine learning applications and reliance on well curated datasets to improve classification outcomes. However, real-world conditions introduce additional complexities that can challenge the accuracy and reliability of these detection algorithms. While studies have explored the behavior of these adventitious sounds in simulated or well-controlled settings [10]–[12], these settings may not accurately represent the dynamic and often noisy conditions encountered in real-world clinical practice. Factors such as background noise, patient state, and diverse comorbidities can significantly influence acoustic signals, complicating data collection and interpretation. Consequently, robust analysis methods and noise-tolerant algorithms are needed to ensure reliable performance under varying clinical circumstances.

One challenge in computerized auscultation analysis is that the broader field of audio signal processing is predominantly driven by speech-related applications, resulting in a reliance on

methodologies originally developed for speech. For example, Mel-Frequency Cepstral Coefficients (MFCCs) —developed for Automatic Speech Recognition and inspired by speech production and perception— have been used for wheeze detection [13] and other time-feature representations were explored for crackle detection [14], [15]. While cepstral analysis can effectively capture the spectral structure of lung sounds, it is effectively tailored to the temporal and spectral characteristics of lung sounds, particularly the focus on resonance (or formant) features in speech production. Multi-resolution methods based on wavelet transform [16]–[19] were also shown to effectively address the detection of abnormal lung sounds. More recent work has posed the adventitious lung recognition problem as a classification task in the source domain [20], [21], leading to more efficient and automated respiratory sound analysis [22].

Recent advances in deep learning and signal processing have significantly improved automated respiratory sound analysis capabilities. Modern approaches leverage sophisticated neural network architectures for feature extraction and classification, demonstrating enhanced accuracy in detecting adventitious lung sounds. CNN-LSTM hybrid models have achieved accuracies of 66.31% for crackle and wheeze classification [23], while the LungPass platform demonstrated high sensitivity (96.9%) and specificity (90%) in identifying normal lung sounds [24]. Further improvements came through depthwise separable CNNs, achieving 85.74% accuracy while maintaining computational efficiency [25]. The field has continued to evolve with multi-task learning approaches, where MobileNet-based models achieved 74% accuracy for lung sound analysis while simultaneously performing disease classification with 91% accuracy [26].

Despite notable progress, existing approaches to respiratory sound classification often constrain themselves to rigid problem formulations that may not translate well to real-world clinical scenarios. A comprehensive analysis by [27] demonstrated that classification performance varies significantly based on how the input data is structured and presented to the models. For instance, models trained on fixed-duration segments show marked degradation when evaluated on variable-length recordings, with accuracy drops of up to 35% in some cases. This disparity highlights a critical gap between controlled research environments and practical clinical applications. Moreover, the same algorithmic approach yields substantially different results when the classification task is reformulated (e.g. input duration, random events), raising concerns about how reliably they would perform under real-world clinical conditions where input data and classification criteria are not as rigidly controlled.

A significant challenge in many machine learning approaches is the rigidity of the structure of training and inference pipelines, which typically demand uniformly sized inputs to feed into models and need for careful annotations of all samples available. In practice, clinicians do not assess respiratory sounds in fixed-duration segments but rather evaluate the entire auscultation recording, paying particular attention to segments that exhibit potential anomalies. They adapt their analysis based on the quality and characteristics of the sounds, considering varying durations and intensities of adventitious events. This natural diagnostic process suggests that automated systems should similarly incorporate flexible, adaptive approaches that can handle variable-duration inputs while maintaining consistent performance across different recording conditions.

Building on previous findings that demonstrate the effectiveness of Variational Recurrent Neural Networks (VRNN) classifiers in creating discriminative feature spaces for auscultation analysis [28] and that confident model predictions correlate strongly with physician annotations [29], we propose a novel multi-stage approach for automated respiratory sound analysis. This approach mirrors the physician diagnostic process by integrating both classical signal processing and deep learning across multiple temporal scales. Specifically, it employs an expert-driven annotation strategy in three core stages: (1) selecting segments of interest from complete auscultation recordings, (2) performing segment-level classification, and (3) adjudicating the final recording-level label based on carefully chosen segment predictions. This design ensures robust performance in realistic clinical contexts, bridging the gap between controlled research environments and practical deployments.

To achieve these objectives, our approach introduces a specialized audio segment selection technique that effectively isolates clinically relevant regions within respiratory recordings. Isolated segments are then evaluated using parallel processing paths tailored for wheeze and crackle detection before integrated assessment of the entire recording is achieved using a balanced random forest classifier. Our approach achieves enhanced sensitivity and specificity compared to existing methods while maintaining clinical interpretability. Key contributions of the current study include: 1) a robust segment selection algorithm achieving over 98% recall in identifying relevant segments, 2) a balanced multi-path classification approach effectively handling class imbalance, and 3) a clinically validated fusion technique that leverages previous insights about confident predictions to enable reliable diagnosis.

II. Data

The auscultation audio data analyzed in this paper is acquired as part of the Pneumonia Etiology Research for Child Health (PERCH) study group [30]. This comprehensive dataset consists of respiratory recordings obtained from 792 pediatric patients aged 1-59 months, representing seven geographically diverse countries (The Gambia, Mali, Kenya, South Africa, Zambia, Bangladesh and Thailand). Lung sound acquisition is performed using a Thinklabs digital stethoscope, with each recording capturing approximately 7-28 seconds of respiratory activity at eight chest positions. The initial data collection protocol involved recording the lung sounds at 44.1 kHz sampling frequency, followed by a series of preprocessing steps for the subsequent computerized auscultation analysis. These steps include low-pass filtering the audio signal using a fourth-order Butterworth filter with a 4 kHz cutoff frequency, downsampling to 8 kHz, and normalizing to achieve zero mean and unit variance. To further refine the audio quality, a spectral subtraction based noise cancellation algorithm is

employed [31]. This algorithm effectively mitigates various sources of signal degradation, including clipping distortions, mechanical and sensor artifacts, cardiac sound interference, intense vocalization from subjects, and ambient environmental noise.

The dataset undergoes a rigorous annotation process involving nine expert reviewers, all of whom are pediatricians or physicians. The reviewers assess the recordings blindly, without access to any additional clinical information, ensuring an objective appraisal of auscultation signals. During the evaluation, reviewers categorize each recording as either a normal auscultation signal or one containing adventitious patterns, specifically identifying the presence of wheezes or crackles. For recordings classified as abnormal, the reviewers precisely indicate the temporal boundaries of an example anomalous region in the recording, marking both the onset and offset of these acoustic irregularities. To enhance annotation reliability, each recording is independently evaluated by two reviewers. In cases where these initial assessments diverge, a third reviewer arbitrates to achieve consensus. Overall, the dataset consists of 14.7 hours of normal recordings and 5.11 hours of abnormal recordings (broken into 3.81 hours containing wheezes, and 1.3 hours with crackles). Among the abnormal recordings, expert physicians have selected 1.9 hours of wheezes and 38 minutes of crackles to represent annotated segments with these labels. .

Data Augmentation: This data contains relatively few instances of crackles resulting in a pronounced class imbalance. To address this issue, we implement a specialized augmentation technique to generate synthetic data, as described in prior work [32]. This approach, inspired by SMOTE (Synthetic Minority Over-sampling Technique), is tailored specifically for respiratory sounds and is applied only to the 2-second segment-level classification part of our pipeline. The augmentation focuses specifically on selected 2-second windows that overlap with regions identified by expert physicians as containing abnormal sounds, thereby generating synthetic training examples for underrepresented classes. Importantly, this augmentation method is utilized only during the training phase of the segment-level classifier, ensuring improved classifier robustness without compromising the integrity of evaluation outcomes. We double the abnormal classes by adding as many synthetic samples as there is original data for both wheezes and crackles.

III. Methods

The proposed method is a novel classification framework designed specifically to analyze entire respiratory recordings obtained from individual chest positions by systematically addressing variability in adventitious sound characteristics. Our approach uniquely incorporates a specialized segment-selection strategy, tailored to identify clinically relevant audio segments that are sensitive to specific adventitious sounds (crackles or wheezes). These segments are then independently processed through parallel classification paths dedicated separately to each type of adventitious sound, generating predictions along with associated confidence scores. The framework further leverages these confidence-informed segment-level classifications by fusing results across all segments, ultimately producing a robust, multi-class decision for the entire lung recording.

A. Anomaly Specific Segment Selection

This initial stage of the proposed method focuses on identifying segments of interest within the overall auscultation recording that best represent wheezes or crackles -if present, serving as foundation for subsequent model inference. Recognizing the distinct acoustic characteristics of these two types of abnormal lung sounds, the approach employs separate selection criteria tailored specifically to wheezes and crackles. These criteria are developed based on an analysis of annotated segment boundaries for wheezes and crackles, particularly focusing on signal amplitudes, density of occurrence in time as well as transience characteristics. Recent work has demonstrated the importance of tailored time-frequency representations such as the Tunable Q-factor Wavelet Transform (TQWT) in enhancing statistical feature extraction of adventitious lung sounds [9].We specifically note that wheezes are typically sustained over longer periods of time with smoother rise time while crackles are far sharper, transient and shorter over time. Figure 1 shows the distribution of distances between the most prominent peaks within the annotated anomalous regions. To perform this analysis, we first compute the analytical signal corresponding to expert-annotated wheezes and crackles by applying a Hilbert transform with a 500 ms window. We then identify the most prominent peaks in the smoothed audio waveform and calculate the distances between them. The results reveal a sharper drop-off in peak-to-peak distances at shorter durations for crackles compared to wheezes, highlighting the more transient nature of crackles. Building on this insight, the proposed technique pinpoints salient time segments within the entire stethoscope recordings that are of particular interest for subsequent normal/abnormal lung sound classification.

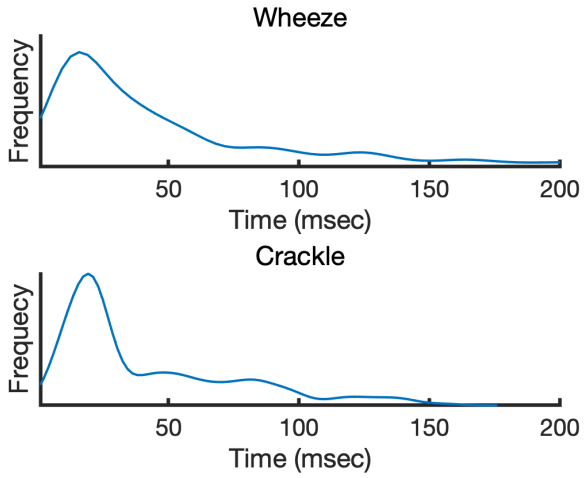

Fig. 1. The distribution of distance between most prominent peaks in annotator selected wheeze and crackle segments.

The process begins by applying a Hilbert transform to

the entire audio recording collected from an individual chest position. The transformed signal is then smoothed and low-pass filtered using a fifth-order Butterworth filter with a cutoff frequency fc. Peaks are then identified in the filtered signal, ensuring they are at least a predefined duration T_d apart and rank within a specified top percentile Pc in terms of peak amplitude. The three hyperparameters for wheeze and crackle detection, as determined from the heuristic analysis for identifying wheezes and crackles - hand-tuned to fit the aggregate wheeze and crackle behavior, are presented in TABLE I. Fig. 3 provides a visual representation of this centering process for wheeze preprocessing.

Baseline: To assess the proposed segment selection approach, a baseline method using a moving 2-second window is implemented, following typical approaches used in most audio systems [33], [34]. This approach traverses the entire recording, providing a standardized reference for comparison with the proposed anomaly-specific segment selection method.

TABLE I
WHEEZE AND CRACKLE SPECIFIC HYPER PARAMETERS FOR SALIENT AUSCULTATION ANOMALY CENTERING PREPROCESSING

Hyperparameter	Wheeze	Crackle
fc **(Hz)**	50	300
T_d **(msec)**	150	25
Pc	70	90

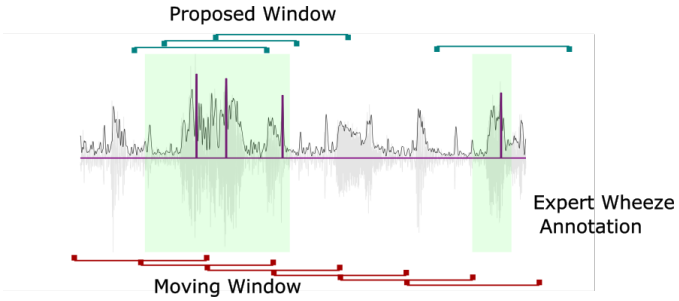

Fig. 2. Example of Wheeze preprocessing with the expert panel onset-offset region highlighted in blue, proposed segment selection shown as green segment above the waveform and baseline moving window shown in red below the waveform.

B. Audio Segment Level Binary Classification

Given the unique profiles of wheeze and crackle sounds, the proposed scheme develops parallel class-specific binary classification with two flows focusing on Wheeze and Crackle sounds respectively. Each flow utilizes an identical classifier architecture, designated as N-W for normal vs. wheeze classification and N-C for normal vs. crackle classification. Each classifier follows a structured pipeline: 2-second audio segments, selected using the anomaly-specific windowing approach described above. The segments are centered around the most representative wheeze and crackle peaks identified through their respective segment-selection method. These segments serve as inputs to the respective binary classification block. During training, wheeze-specific and crackle-specific segments are labeled accordingly for their respective classifiers. For class assignment, all segments extracted from normal recordings are labeled as normal, while only windows overlapping with manually annotated onset-offset regions in abnormal recordings are labeled as abnormal for training the segment-level classifiers.

Auscultation audio signals are converted to mel-frequency spectrograms of dimensions 63x64 with nfft=512, 50% window overlap, and 64 nmels using librosa [35]. It essentially transforms the temporal signal to a spectrogram mapped onto a bank of log-scaled asymmetrical cochlear filters. The melspectrograms are then processed using the architecture described below.

Contrastive VRNN Binary Classifier: The proposed architecture operates in a stochastically enhanced feature space designed to improve the separation between normal and abnormal lung sounds, followed by a fully connected neural network (FC) for final classification. A variational recurrent neural network (VRNN) serves as feature extractor , mapping the spectro-temporal representations of audio signals to a contrastive stochastic recurrent feature space. The input mel-spectrograms of dimensions $t_{len} \times f$ are sequentially encoded into a feature space of size $t_{len} \times z$ using a recurrent temporal encoder with $z_{dim} = 4$ and 2 hidden layers. These stochastic features are then passed through an FC network with two hidden layers. The first layer maps the feature to an embedding space e with $e_{dim} = 8$ and the second layer maps the embedding to a logit space. This binary classifier processes temporal dependencies in the stochastic feature space and assigns each segment to either the normal or abnormal class. A final sigmoid activation function produces confidence scores for each prediction.

This network is trained using a combination of two loss functions

- Contrastive Loss - a supervised loss that maximizes separation between normal and abnormal classes in the stochastic recurrent feature space ($t_{len} \times z$),
- Cross-entropy loss - applied to the binary labels and logits outputs from the classifier, where "0" represents normal lung sounds and "1" represents abnormal sounds (wheezes or crackles).

The model is optimized using Adam optimizer with learning rate of 0.001 over 30 epochs, with a weighted combination of contrastive and cross entropy loss. Performance is conducted using 5-fold cross-validation to ensure robustness.

Baseline: To ensure a robust evaluation, we compare our approach against recently published state-of-the-art baseline systems from the literature, which follow a similar framework of fixed window-size inputs and binary classification of normal vs. abnormal lung sounds. These baseline models serve as reference points, allowing us to assess the effectiveness of classification branches. The chosen baselines are:

1) CNN-LSTM model [36] in which the Convolutional Neural Network feature extractor captures the short-range dependencies using local receptive fields and a bidirectional LSTM classifier sequentially decodes the feature to give a final classification
2) Multi-View Spectrogram Transformer (MVST) [37] that splits the input melspectrogram into different size

patches that are fed into transformer encoders to extract self-attention features across multi-view acoustic elements to give the final classification

3) Multi-branch Temporal Convolution Network integrated with a Squeeze-and-Excitation (SE) Network (MBTC-NSE) [38] where three TCN branches capture information of the input melspectrogram at different receptive fields. The 3D representation is then parsed by a Conv2D block with SE block - a component that strengthens the CNN representations through dynamic channel-wise feature calibration to give the final classification.

For fair comparison, we standardize model complexities across all implementations to approximately 10K parameters by modifying model architecture sizes and audio input dimensions. This adjustment is necessary not only for fairness but also to prevent overfitting, as we observed that more complex models failed to generalize well given the limited dataset size. The models are then trained as described in the original works on the normal and abnormal segments obtained in a moving window fashion. Furthermore, while the original comparative baseline implementations did not mention any data augmentation or imbalance handling techniques, we applied weighted resampling during training for all models to address class imbalance issues.

C. Site-Level Audio Tagging

The ultimate objective is to tag an entire audio recording as either normal, wheeze, or crackle auscultation. To achieve this, we integrate the two previously developed blocks using a fusion block. Each classifier branch results in a sequence of posteriors from the N-W and N-C classifiers corresponding to different segments extracted from the audio recording along each branch. These posteriors are then aggregated using various strategies. Our proposed data-driven approach optimizes the fusion strategy in a supervised manner. A rule-based strategy is also explored as alternative baseline approach.

1) Data-Driven Fusion: We implement a data-driven fusion system that integrates the posterior probabilities from the two parallel classifiers (N-W for wheeze detection and N-C for crackle detection). Instead of relying on simple heuristic-based aggregation, we employ a supervised fusion model using a Random Forest classifier. A fixed length (10) selection of posteriors from both classifiers is concatenated and given as input to 3-way classifier that ultimately yields a class label for the entire audio recording.

Random Forest Classifier: A balanced random forest classifier is designed with 20 input nodes and 3 output nodes. This ensemble method combines multiple decision trees, each trained on a class-balanced subset of the data to address class imbalance. During training, individual trees are constructed using the Gini impurity criterion to determine optimal splits, maximizing class separation at each node. The classifier is optimized by exploring various hyperparameters such as the number of estimators, the maximum depth, and the feature selection strategies. The model is trained using the recording-level labels of the stethoscope recordings. Final predictions are made through majority voting across all trees to enhance the classification accuracy and robustness.

Posterior Selection: We employ two selection criterion on the 2-second posterior outputs of the N-W and N-C classifiers.

- *Most Confident Posteriors*: Select the 10 most confident posterior probabilities (farthest from 0.5) from each branch.
- *Random Selection*: Randomly select 10 posterior probabilities from each branch (wheeze and crackle). This selection approach is employed as baseline to evaluate the informative nature of the posteriors in terms of classifier confidence.

2) Rule-Based Fusion: An alternative fusion system is to use rule-based decision framework to fuse posteriors from the N-W, N-C classifier branches.

The simplest baseline involves reducing posterior probabilities from the 2-second segments for each pipeline to a single probability value and applying a rule-based decision framework. Specifically:

TABLE II
CLASSIFICATION RULES BASED ON POSTERIOR PROBABILITIES

Condition	Classification
$P(\text{Wheeze}) < 0.5$ and $P(\text{Crackle}) < 0.5$	Normal
$P(\text{Wheeze}) > P(\text{Crackle})$	Wheeze
$P(\text{Wheeze}) < P(\text{Crackle})$	Crackle

The rule-based fusion involves no training and hence can be directly applied at inference following the Segment-Level Classifier. We apply two types of posterior reductions: a) Aggregate Rule-Based Decision: Average the posteriors for each class; b) Confident Rule-Based Decision: The probability value farthest away from 0.5 serves as the most-confident segment prediction in a particular recording.

IV. RESULTS

We analyze the performance of the proposed system at all three stages: Segment selection, segment-level binary classification, and recording-level audio tagging. Throughout all experiments, the train-validation-test splits are kept consistent across all levels to ensure fairness in evaluation across methods. By systematically comparing these approaches, we demonstrate that leveraging the most confident posteriors with a balanced Random Forest with the proposed classification technique yields superior classification performance for tagging auscultations at the recording level.

A. Segment Selection

The proposed segment selection technique aims to identify regions of interest within the entire recording. Its effectiveness is evaluated against expert physician annotations. We define the accuracy of the segment-selection algorithm as the proportion of expert-selected segments that overlap with those identified by our algorithm.

For crackle detection, our method achieves a recall of $98.16\% \pm 1.1\%$, while for wheeze detection, it attains a recall of $98.35\% \pm 1.07\%$. These high recall values indicate that our preprocessing algorithm successfully captures the vast majority of regions identified as significant by expert

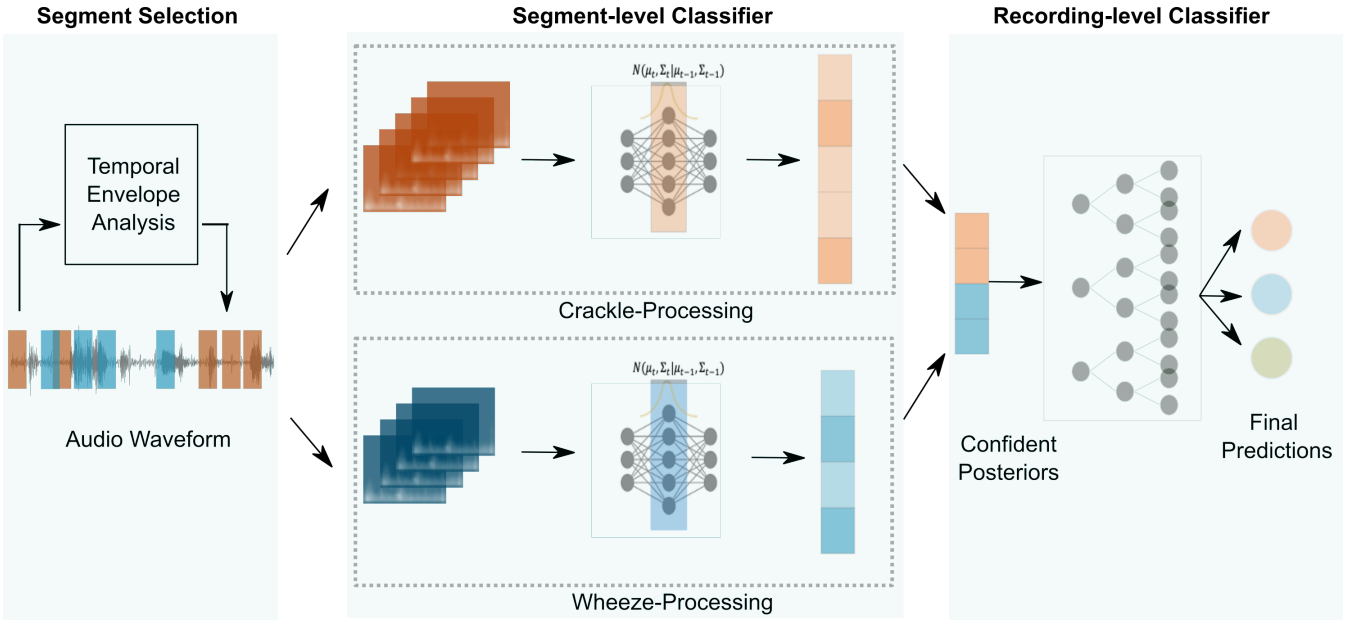

Fig. 3. Overview of methods at inference: Every audio recording is first assessed for potential wheeze and crackle segments at the segment selection level. Each of the selected segments are processed by corresponding anomaly-specific segment-level classifier to assign probability of wheezes and crackles for respective candidate segments. Adjudication at recording level is obtain by selecting the most confident predictions from each anomaly-processing pipeline and inferred by a trained balanced-random forest to give the final three-way selection.

physicians. It is important to note that in recordings annotated as abnormal lung sounds, physicians typically highlight the most representative adventitious durations. However, this does not preclude the existence of abnormal breathing patterns outside their selection. Consequently, we prioritize recall over precision in our evaluation, as there is no definitive ground truth for true negatives in this context. As noted in Methods, this approach is contrasted with a running two-second window with 50% overlap. Ultimately, the final evaluation at the classification stage provides an accurate evaluation of the value of the proposed segment selection.

B. Segment Level classification

Table III presents the segment-level classification performance of our proposed method alongside baseline comparison systems. The performance is reported as specificity and sensitivity for both wheeze and crackle classification (focusing on each branch of the system), as well as the combined AUC ROC. The proposed system is evaluated with two variations: one employing the proposed segment selection and one using a moving window with the same classifier. In addition, we compare these systems against three existing baselines: CNN-LSTM [36], MVST [37], and TCN [38] (see III for details).

Among the baseline models, CNN-LSTM exhibits high specificity but struggles with sensitivity, especially for crackle detection. This indicates that while the model effectively minimizes false positives (specificity $> 85\%$), it frequently fails to detect actual crackle segments, leading to a high false negative rate. In contrast, MVST and TCN show more balanced performance between specificity and sensitivity. These models maintain a better ability to detect both normal and abnormal segments, reducing the risk of underdiagnosing adventitious lung sounds, particularly crackles. The moving window approach combined with with the proposed classifier, demonstrates improved sensitivity compared to baseline methods, validating the effectiveness of the classifier architecture even without specialized segment selection. Ultimately, the combination of the proposed segment selection method and classifier yields the best overall performance for both wheeze and crackle detection. In wheeze detection, this approach achieves a significant improvement in sensitivity (72.75±3.50%), however with a trade-off in specificity. For crackle detection, the proposed method outperforms baselines in sensitivity (63.6±2.8%) while maintaining competitive specificity levels. A t-test comparing the proposed system to each of the other baseline models confirms that the improvement in the average sensitivity is statistically significant (p-values CNN-LSTM:1.4e-06, MVST:1.3e-04, TCN: 2.7e-04, Moving+Proposed:7.4e-04).

The combined AUC ROC scores further validate the effectiveness of our proposed approach. The full proposed method, incorporating both segment selection and dual-branch classifier, achieves the highest score of 0.77±0.02, indicating improved discrimination across both wheeze and crackle detection tasks. These results highlight the effectiveness of our proposed segment selection method in capturing relevant acoustic features for respiratory anomaly detection, as well as the robustness of our classifier in leveraging these features for accurate classification.

C. Audio Tagging

The final stage of our analysis focuses on audio tagging, where we classify variable-duration chest position recordings as normal, wheeze, or crackle. This stage is crucial as it provides a holistic assessment of respiratory sounds by integrating information from multiple representative 2-second segments within the recording.

TABLE III
RESULTS AT THE SEGMENT-LEVEL CLASSIFICATION

Abnormal Class	Wheeze		Crackle		Combined
Method	Specificity	Sensitivity	Specificity	Sensitivity	AUC ROC
CNN-LSTM [36]	85.6±0.67	50.8±1.71	90.8±0.86	25.4±2.37	0.69±0.01
MVST [37]	83.4±0.67	49.4±1.40	82±1.20	41.8±2.39	0.70±0.01
TCN [38]	78.4±1.94	54.6±3.43	78.4±3.60	44.4±5.07	0.73±0.05
Moving window + Proposed Classifier	86.6±0.67	58.2±2.10	80.6±2.31	50±3.8	0.75±0.02
Proposed segment + Proposed Classifier	71.5±1.29	**72.75±3.50**	72.6±3.2	**63.6±2.8**	**0.77±0.02**

TABLE IV
RECORDING-LEVEL FUSION VALIDATION

Method	Normal	Wheeze	Crackle	Balanced Accuracy
Aggregate Rule-Based Decision	70.19 ± 0.04	51.45 ± 0.03	36.72 ± 0.15	52.78 ± 0.03
Confident Rule-Based Decision	63.26 ± 0.07	63.40 ± 0.03	34.90 ± 0.11	53.83 ± 0.02
Random → Balanced Random Forest	58.58 ± 0.35	59.31 ± 0.99	45.45 ± 1.45	54.4 ± 0.98
Confident → Balanced Random Forest	68.56 ± 1.46	58.95 ± 1.12	58.55 ± 3.30	**62.12 ± 1.22**

Our results, presented in Table IV, demonstrate the superiority of confidence-informed aggregation over naive approaches. The simple aggregate decision method, which essentially averages the predictions across all segments, achieves moderate performance but fails to capture the nuanced information present in the most informative segments. The normal auscultations benefit the most from this method as expected with a specificity if 70.19 ± 0.04. However, given abnormal events occur intermittently within regular breathing patterns, the naive method tends to dilute their impact, leading to a "wash out" effect. This is particularly evident in the lower performance for crackle detection (36.72 ± 0.15 %). For reference, chance level for a 3-way classifier is 33.33%.

A similar rule-based prediction analysis is performed using the most confident posterior instead of the aggregate posterior. While this strategy improves wheeze detection, increasing recall from 51.45 ± 0.03 % to 63.40 ± 0.03 %, it negatively impacts crackle classification. The most confident posterior often corresponds to prominent normal breathing segments, leading to a drop in crackle accuracy from 36.72± 0.15 % in the aggregate rule-based baseline to 34.90 ± 0.11 % in the most-confident rule-based baseline. Additionally, this method reduces specificity from 70.19 ± 0.04 to 63.26 ± 0.07, further highlighting its limitations in capturing transient abnormalities.

To highlight the significance of ensemble aggregation over reducing segment posteriors to a single value per recording, we analyze their density distributions in Fig 4. The first column represents the wheeze pipeline while the second column focuses on the crackle pipeline, with each row corresponding to one of the auscultation classes (Normal, Wheeze, & Crackle). For the abnormal classes, each graph contains two overlaid density plots:

- Red density plot: Posterior probabilities of segments marked abnormal by expert listeners
- Black density plot: Posterior probabilities of segments outside the physician-marked abnormal regions.

The probability values range from 0 (confident normal prediction) to 1 (confident abnormal prediction). Our analysis reveals that for both the wheeze and crackle pipelines, normal class posteriors predominantly cluster near 0. However, within the wheeze and crackle classes, we observe distinct peaks for the abnormal segments, confirming their informative nature (Figure 4row2-left and row3-right, respectively). Interestingly, slight peaks near 1 appear for unmarked segments, reinforcing our earlier observation: Physicians tend to mark the most representative adventitious pattern, but adventitious breathing patterns may exist outside the annotated regions in an abnormal recording.

Building on these observations, our overall system moves beyond single-value reduction for overall classification. We evaluate an ensemble-technique that functions like a voting mechanism, preserving adventitious breathing patterns. Training a balanced random forest on a random selection of posteriors yields a boost in crackle accuracy (45.45 ± 1.45 %) compared to naive baselines, though at the expense of normal and wheeze performance (Table IV).

Rather than relying on random posterior selection, our proposed method leverages the confidence trends of posteriors

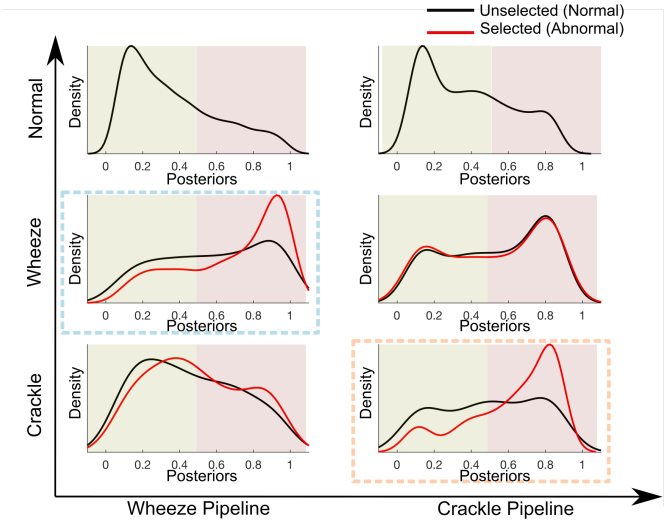

Fig. 4. Probability Outputs from segment-level classifier: The posterior probabilities outputted by the proposed segment-level classifier are plotted as density plots with column one depicting the Wheeze Pipeline and column two depicting the Crackle Pipeline

(Fig. 4) to develop a confidence-based fusion system that ensures most informative segments contribute more effectively to the final classification decision. This system picks the most confident posteriors (farthest away from the decision boundary) among all the segments in each pipeline, and uses these values to train a balanced random forest. This method achieves the best performance across all classes - balanced accuracy (62.12 ± 1.22). A t-test comparing the segment-level average sensitivity with the comparative analysis is performed and the proposed method yields significantly higher average abnormal accuracies (p-values Aggregate Rule-Based:1.34e-09, Confident Rule-Based:1.4e-09, Random $\rightarrow$ Balanced Random Forest :0.014).

V. Discussion

This study introduces a confidence-driven, multi-stage framework for automated detection of wheezes and crackles in pediatric auscultation recordings. Our approach deviates from traditional audio classification systems by structuring the modeling scheme around the clinical reasoning process: First, the approach identifies potentially abnormal regions. Next, it classifies these segments using tailored anomaly-specific classifiers. Finally, it fuses predictions based on model confidence. Compared to existing systems, this framework shows notable improvements in detecting both wheezes and crackles; which are considered two widely encountered trademarks of adventitious auscultations reflecting abnormal lung physiology. Prior models, such as CNN-LSTM hybrids [36] or transformer-based MVST architectures [37], have been shown to often exhibit imbalanced performance, typically favoring one anomaly type (e.g., wheeze) at the cost of another. In contrast, our method achieves balanced performance with improved sensitivity across both categories, highlighting the value of anomaly-specific segmentation and classification.

Moreover, while previous work often relies on fixed-window input sampling [5], [39], the proposed approach employs a dynamic segment selection strategy tailored to the temporal characteristics of each anomaly. Uniform sampling is more commonly encountered in audio analysis systems that analyze signals such as speech and music. However, in the case of auscultation, an adaptive segment-selection mirrors clinical practice, where physicians focus on salient portions of a recording rather than evaluating signals uniformly. Our segment selection achieves over 98% recall against expert annotations, validating its ability to capture clinically meaningful regions. Importantly, this method generalizes across variable-length recordings, addressing limitations seen in earlier models that degrade on out-of-distribution inputs [27]. Other work has shown that refined and tailored signal processing techniques such as wavelet-based methods can distinguish wheeze types with high performance, leveraging tunable time–scale representations [40].

The fusion stage is also a unique element of the proposed approach. Rather than using a simple average of predictions, we employ a confidence-based strategy that emphasizes reliable, discriminative segments. It also mimics an expert panel's adjudication process rather than relying on simple aggregation. This design increases classification robustness, especially for transient anomalies like crackles, which may be diluted in average-based fusion. This approach significantly enhances classification reliability and robustness, reinforcing the importance of ensemble-based decision strategies in biomedical signal processing [5].

Overall, this work advances the clinical translation of automated respiratory sound analysis, particularly for pediatric care in resource-limited settings. While the system outperforms prior work and lays a strong foundation for translation, key challenges remain. Clinical deployment requires generalization across devices and patient populations, necessitating evaluation on additional datasets to assess robustness and transferability. Fine-tuning pretrained deep networks has been shown to be an effective strategy in low-data biomedical settings [41]. Another component of real-world deployment is hardware integration. Current models are trained on high-fidelity digital stethoscope recordings, but in practice, device variability (including differences in frequency response, noise profiles, and microphone sensitivity) can significantly impact model performance. Clinical translation will require validating the system across diverse stethoscope types and clinical settings. Additionally, integration into embedded systems or edge devices presents constraints on memory, power, and compute, which must be addressed through model compression, pruning, or real-time inference optimization. Our modular framework is designed to support such adaptations by decoupling core components, enabling flexibility in how signal processing and classification are deployed across hardware platforms. Beyond technical deployment, further evaluation, algorithmic refinement, and integration into clinical workflows are essential. Regulatory approval and data governance frameworks will also shape deployment pathways. Despite these challenges, this study significantly narrows the translational gap by aligning system design with clinical reasoning and providing a scalable foundation for future prospective trials.

VI. Conclusion

Computerized respiratory sound analysis has traditionally been approached as a multi-label classification task, where performance is largely dictated by how the machine learning framework is designed [39]. However, clinical auscultation relies on a more nuanced decision-making process, integrating expert knowledge and contextual cues [7]. To bridge this gap, we present a multi-stage, confidence-driven framework for wheeze and crackle detection, inspired by physician assessment strategies. Compared to recent state-of-the-art methods [36]–[38], our approach improves sensitivity for both wheeze and crackle detection while maintaining robust performance across variable-length recordings and challenging class imbalance scenarios. The use of contrastive variational recurrent neural networks (CVRNN) enhances discriminative feature learning, while confidence-guided aggregation avoids the pitfalls of naive averaging commonly seen in existing systems.

Our approach improves on recent works [36]–[38] by incorporating anomaly-specific segment selection, segment-level classification, and an intelligent fusion strategy for final audio tagging.

Importantly, our method aligns machine learning architecture design with the practical realities of auscultation assessment, offering a more interpretable and adaptable solution. These strengths make it particularly well-suited for deployment in low-resource settings where diagnostic tools and clinical expertise may be limited.

Future work will focus on expanding validation across broader populations and integrating complementary physiological signals including multimodal lung diagnostics to improve contextual understanding [42]. Additionally, leveraging self-supervised learning and domain adaptation may enhance model generalization in real-world clinical environments. Ultimately, our framework moves toward more reliable, explainable, and deployable computer-aided diagnostic tools for pediatric respiratory care.

While it is common for audio analysis systems (speech, music, auscultations) to rely on uniform sampling of signal segments across time [33], [34], [43], our method introduces adaptive segment selection tailored to wheeze- and crackle-specific acoustic features. This approach aligns with findings from clinical studies that emphasize the importance of focusing on localized, representative adventitious patterns rather than treating respiratory cycles as uniform entities. Additionally, the proposed contrastive variational recurrent neural classifier (CVRNN) improves feature extraction for robust classification, overcoming the common limitation of low sensitivity in adventitious sound detection reported in previous works [39], [43].

The final confidence-driven fusion leverages posterior probabilities to refine decision-making, mimicking an expert panel's adjudication process rather than relying on simple aggregation. This approach significantly enhances classification reliability and robustness, reinforcing the importance of ensemble-based decision strategies in biomedical signal processing [5]. By integrating machine learning innovations with expert-inspired strategies, our work advances automated respiratory sound classification towards greater clinical applicability. The improved robustness and interpretability of our method hold promise for early disease detection and remote respiratory monitoring, particularly for complex lower respiratory conditions like pneumonia. Future research should explore self-supervised learning, domain adaptation, and multimodal fusion with additional physiological signals to further enhance generalizability and real-world deployment.

REFERENCES

[1] D. Marangu and H. J. Zar, "Childhood pneumonia in low-and-middle-income countries: An update," *Paediatric Respiratory Reviews*, vol. 32, pp. 3–9, 11 2019.

[2] D. A. McAllister, L. Liu, T. Shi, Y. Chu, C. Reed, J. Burrows, D. Adeloye, I. Rudan, R. E. Black, H. Campbell, and H. Nair, "Global, regional, and national estimates of pneumonia morbidity and mortality in children younger than 5 years between 2000 and 2015: a systematic analysis," *The Lancet Global Health*, vol. 7, no. 1, pp. e47–e57, 1 2019.

[3] E. D. McCollum, D. E. Park, N. L. Watson, N. S. S. Fancourt, C. Focht, H. C. Baggett, W. A. Brooks, S. R. C. Howie, K. L. Kotloff, O. S. Levine, S. A. Madhi, D. R. Murdoch, J. A. G. Scott, D. M. Thea, J. O. Awori, J. Chipeta, S. Chuananon, A. N. DeLuca, A. J. Driscoll, B. E. Ebruke, M. Elhilali, D. Emmanouilidou, L. P. Githua, M. M. Higdon, L. Hossain, Y. Jahan, R. A. Karron, J. Kyalo, D. P. Moore, J. M. Mulindwa, S. Naorat, C. Prosperi, C. Verwey, J. E. West, M. D. Knoll, K. L. O'Brien, D. R. Feikin, and L. L. Hammitt, "Digital auscultation in PERCH: Associations with chest radiography and pneumonia mortality in children," *Pediatric Pulmonology*, vol. 55, no. 11, pp. 3197–3208, 11 2020.

[4] D. Buonsenso and C. De Rose, "Implementation of lung ultrasound in low- to middle-income countries: a new challenge global health?" *European Journal of Pediatrics*, vol. 181, no. 1, pp. 1–8, 1 2022.

[5] P. Kapetanidis, F. Kalioras, C. Tsakonas, P. Tzamalis, G. Kontogiannis, T. Karamanidou, T. G. Stavropoulos, and S. Nikoletseas, "Respiratory Diseases Diagnosis Using Audio Analysis and Artificial Intelligence: A Systematic Review," *Sensors*, vol. 24, no. 4, p. 1173, 2 2024.

[6] R. L. Wilkins, J. R. Dexter, R. L. H. Murphy, and E. A. DelBono, "Lung Sound Nomenclature Survey," *Chest*, vol. 98, no. 4, pp. 886–889, 1990.

[7] Z. Moussavi, *Fundamentals of respiratory sounds and analysis*, 1st ed. Morgan & Claypool Publishers, 2007.

[8] L. J. Hadjileontiadis, *Lung Sounds: An Advanced Signal Processing Perspective*. San Rafael, California: Morgan & Claypool Publishers, 1 2008, vol. 3, no. 1.

[9] B. Cansiz, C. U. Kilinc, and G. Serbes, "Tunable Q-factor wavelet transform based lung signal decomposition and statistical feature extraction for effective lung disease classification," *Computers in Biology and Medicine*, vol. 178, p. 108698, 8 2024.

[10] N. Q. Al-Naggar, "A new method of lung sounds filtering using modulated least mean square—Adaptive noise cancellation," *Journal of Biomedical Science and Engineering*, vol. 6, pp. 869–876, 2013.

[11] K. K. Guntupalli, P. M. Alapat, V. D. Bandi, and I. Kushnir, "Validation of Automatic Wheeze Detection in Patients with Obstructed Airways and in Healthy Subjects," *Journal of Asthma*, vol. 45, no. 10, pp. 903–907, 2008.

[12] J. Li and Y. Hong, "Wheeze Detection Algorithm Based on Spectrogram Analysis," in *2015 8th International Symposium on Computational Intelligence and Design (ISCID)*, vol. 1. IEEE, 12 2015, pp. 318–322.

[13] J.-C. Chien, H.-D. Wu, F.-C. Chong, and C.-I. Li, "Wheeze Detection Using Cepstral Analysis in Gaussian Mixture Models," in *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, 2007, pp. 3168–3171.

[14] G. Serbes, C. O. Sakar, Y. P. Kahya, and N. Aydin, "Feature extraction using time-frequency/scale analysis and ensemble of feature sets for crackle detection," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pp. 3314–3317, 2011.

[15] ——, "Pulmonary crackle detection using time–frequency and time–scale analysis," *Digital Signal Processing*, vol. 23, no. 3, pp. 1012–1021, 5 2013.

[16] D. Emmanouilidou, K. Patil, J. West, and M. Elhilali, "A multiresolution analysis for detection of abnormal lung sounds," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2012. IEEE, 8 2012, pp. 3139–3142.

[17] S. Ulukaya, G. Serbes, and Y. P. Kahya, "Overcomplete discrete wavelet transform based respiratory sound discrimination with feature and decision level fusion," *Biomedical Signal Processing and Control*, vol. 38, pp. 322–336, 9 2017.

[18] S. Ulukaya, G. Serbes, I. Sen, and Y. P. Kahya, "A lung sound classification system based on the rational dilation wavelet transform," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, vol. 2016-Octob. IEEE, 8 2016, pp. 3745–3748.

[19] S. Ulukaya, G. Serbes, and Y. P. Kahya, "Resonance based separation and energy based classification of lung sounds using tunable wavelet transform," *Computers in Biology and Medicine*, vol. 131, p. 104288, 4 2021.

[20] Q. T. Do, K. Lipatov, H. Y. Wang, B. W. Pickering, and V. Herasevich, "Classification of Respiratory Conditions using Auscultation Sound," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pp. 1942–1945, 2021.

[21] J. Oliveira, F. Renna, P. D. Costa, M. Nogueira, C. Oliveira, C. Ferreira, A. Jorge, S. Mattos, T. Hatem, T. Tavares, A. Elola, A. B. Rad, R. Sameni, G. D. Clifford, and M. T. Coimbra, "The CirCor DigiScope Dataset: From Murmur Detection to Murmur Classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 6, pp. 2524–2535, 6 2022.

[22] Y. Kim, Y. K. Hyon, S. Lee, S. D. Woo, T. Ha, and C. Chung, "The coming era of a new auscultation system for analyzing respiratory sounds," *BMC Pulmonary Medicine*, vol. 22, no. 1, pp. 1–11, 12 2022.

[23] S. Im, T. Kim, C. Min, S. Kang, Y. Roh, C. Kim, M. Kim, S. H. Kim, K. Shim, J.-s. Koh, S. Han, J. Lee, D. Kim, D. Kang, and S. Seo, "Real-

time counting of wheezing events from lung sounds using deep learning algorithms: Implications for disease prediction and early intervention," *PLOS ONE*, vol. 18, no. 11, p. e0294447, 11 2023.

[24] E. A. Lapteva, O. N. Kharevich, V. V. Khatsko, N. A. Voronova, M. V. Chamko, I. V. Bezruchko, E. I. Katibnikova, E. I. Loban, M. M. Mouawie, H. Binetskaya, S. Aleshkevich, A. Karankevich, V. Dubinetski, J. Vestbo, and A. G. Mathioudakis, "Automated lung sound analysis using the LungPass platform: a sensitive and specific tool for identifying lower respiratory tract involvement in COVID-19," *The European Respiratory Journal*, vol. 58, no. 6, p. 2101907, 12 2021.

[25] 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*, vol. 11, no. 4, p. 732, 4 2021.

[26] S. K, D. Koppad, P. Kumar, N. A. Kantikar, and S. Ramesh, "Multi-Task Learning for Lung sound & Lung disease classification," 4 2024.

[27] B. M. Rocha, D. Filos, L. Mendes, G. Serbes, S. Ulukaya, Y. P. Kahya, N. Jakovljevic, T. L. Turukalo, I. M. Vogiatzis, E. Perantoni, E. Kaimakamis, P. Natsiavas, A. Oliveira, C. Jácome, A. Marques, N. Maglaveras, R. Pedro Paiva, I. Chouvarda, and P. de Carvalho, "An open access database for the evaluation of respiratory sound classification algorithms," *Physiological Measurement*, vol. 40, no. 3, p. 035001, 3 2019.

[28] A. Kala and M. Elhilali, "Robust Anomaly Detection of Adventitious Auscultation Signals using Bayesian Belief Tracking," in *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 7 2024, pp. 1–5.

[29] A. Kala, E. D. McCollum, and M. Elhilali, "Implications of clinical variability on computer-aided lung auscultation classification," in *44th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 7 2022, pp. 4421–4425.

[30] E. McCollum, D. Park, N. Watson, C. Focht, C. Bunthi, B. Ebruke, M. Elhilali, D. Emmnouilidou, L. Hossain, D. Moore, A. Mudaua, J. Mulindwa, J. West, K. O'Brien, D. Feikin, and L. Hammitt, "Digitally-recorded lung sounds and mortality among children 1-59 months old with pneumonia in the Pneumonia Etiology research for Child Health study," Tech. Rep., 2017.

[31] D. Emmanouilidou, E. D. McCollum, D. E. Park, and M. Elhilali, "Computerized Lung Sound Screening for Pediatric Auscultation in Noisy Field Environments," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 7, pp. 1564–1574, 2018.

[32] A. Kala and M. Elhilali, "Constrained Synthetic Sampling for Augmentation of Crackle Lung Sounds," in *45th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 7 2023, pp. 1–5.

[33] L. Rabiner and B. Juang, *Fundamentals of Speech Recognition*. Prentice Hall, 1993.

[34] N. Collins, *Introduction to Computer Music*, 1st ed. John Wiley and Sons, 2009.

[35] B. Mcfee, C. Raffel, D. Liang, D. P. W. Ellis, M. Mcvicar, E. Battenberg, and O. Nieto, "librosa: Audio and Music Signal Analysis in Python," *PROC. OF THE 14th PYTHON IN SCIENCE CONF*, 2015.

[36] P. Zhang, A. Swaminathan, and A. A. Uddin, "Pulmonary disease detection and classification in patient respiratory audio files using long short-term memory neural networks," *Frontiers in Medicine*, vol. 10, p. 1269784, 11 2023.

[37] W. He, Y. Yan, J. Ren, R. Bai, and X. Jiang, "Multi-View Spectrogram Transformer for Respiratory Sound Classification," in *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 4 2024, pp. 8626–8630.

[38] Z. Zhao, Z. Gong, M. Niu, J. Ma, H. Wang, Z. Zhang, and Y. Li, "Automatic Respiratory Sound Classification Via Multi-Branch Temporal Convolutional Network," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 2022-May. IEEE, 5 2022, pp. 9102–9106.

[39] B. M. Rocha, D. Pessoa, A. Marques, P. Carvalho, and R. P. Paiva, "Automatic Classification of Adventitious Respiratory Sounds: A (Un)Solved Problem?" *Sensors*, vol. 21, no. 1, p. 57, 12 2020.

[40] S. Ulukaya, G. Serbes, and Y. P. Kahya, "Wheeze type classification using non-dyadic wavelet transform based optimal energy ratio technique," *Computers in Biology and Medicine*, vol. 104, pp. 175–182, 1 2019.

[41] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A Survey on Deep Transfer Learning," 8 2018.

[42] B. Cansiz, C. U. Kilinc, and G. Serbes, "Deep learning-driven feature engineering for lung disease classification through electrical impedance tomography imaging," *Biomedical Signal Processing and Control*, vol. 100, p. 107124, 2 2025.

[43] R. X. A. Pramono, S. Bowyer, and E. Rodriguez-Villegas, *Automatic adventitious respiratory sound analysis: A systematic review*, 2017, vol. 12, no. 5.