electronics

MDPI

Review

Advances and Challenges in Respiratory Sound Analysis: A Technique Review Based on the ICBHI2017 Database

Shaode Yu [1], Jieyang Yu [1], Lijun Chen [1], Bing Zhu [1,*], Xiaokun Liang [2], Yaoqin Xie [2] and Qiurui Sun [3,*]

[1] School of Information and Communication Engineering, Communication University of China, Beijing 100024, China; yushaodecuc@cuc.edu.cn (S.Y.); 202420085410012@mails.cuc.edu.cn (J.Y.); chenlijun@mails.cuc.edu.cn (L.C.)

[2] Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518000, China; xk.liang@siat.ac.cn (X.L.); yq.xie@siat.ac.cn (Y.X.)

[3] Center of Information & Network Technology, Beijing Normal University, Beijing 100875, China

* Correspondence: zhubing1218@cuc.edu.cn (B.Z.); qiuruisun@bnu.edu.cn (Q.S.)

Abstract

Respiratory diseases present significant global health challenges. Recent advances in respiratory sound analysis (RSA) have shown great potential for automated disease diagnosis and patient management. The International Conference on Biomedical and Health Informatics 2017 (ICBHI2017) database stands as one of the most authoritative open-access RSA datasets. This review systematically examines 135 technical publications utilizing the database, and a comprehensive and timely summary of RSA methodologies is offered for researchers and practitioners in this field. Specifically, this review covers signal processing techniques including data resampling, augmentation, normalization, and filtering; feature extraction approaches spanning time-domain, frequency-domain, joint time–frequency analysis, and deep feature representation from pre-trained models; and classification methods for adventitious sound (AS) categorization and pathological state (PS) recognition. Current achievements for AS and PS classification are summarized across studies using official and custom data splits. Despite promising technique advancements, several challenges remain unresolved. These include a severe class imbalance in the dataset, limited exploration of advanced data augmentation techniques and foundation models, a lack of model interpretability, and insufficient generalization studies across clinical settings. Future directions involve multi-modal data fusion, the development of standardized processing workflows, interpretable artificial intelligence, and integration with broader clinical data sources to enhance diagnostic performance and clinical applicability.

Keywords: respiratory sound analysis; the ICBHI2017 database; machine learning; deep learning; abnormal sound categorization; pathological state recognition

check for updates

Academic Editor: Yuankai Qi

Received: 17 June 2025

Revised: 3 July 2025

Accepted: 8 July 2025

Published: 11 July 2025

Citation: Yu, S.; Yu, J.; Chen, L.; Zhu, B.; Liang, X.; Xie, Y.; Sun, Q. Advances and Challenges in Respiratory Sound Analysis: A Technique Review Based on the ICBHI2017 Database. *Electronics* **2025**, *14*, 2794. https://doi.org/10.3390/electronics14142794

1. Introduction

Respiratory diseases, including asthma, chronic obstructive pulmonary disease (COPD), lung cancer, and tuberculosis, are among the leading causes of death globally [1]. The burden of these diseases has been exacerbated by the coronavirus disease 2019, particularly in developing regions with limited access to healthcare services [2].

Respiratory sounds play a critical role in the respiratory and related disease analysis. These sounds are generated by vibrations in the airways and provide valuable information about airway conditions and lung function. Abnormal sounds, such as wheezes, crackles, and stridor, can signal the presence of specific pathological conditions. Moreover, the acquisition of respiratory sounds is noninvasive and cost-effective, so they are especially useful in resource-limited settings. By capturing these acoustic signals, clinicians can assess disease severity, monitor progression in real time, enable early detection, and evaluate treatment efficacy [3].

1.1. Review Studies on Respiratory Sound Analysis

To ensure the reliable diagnosis of respiratory diseases, the development of objective RSA techniques has gained increasing attention. Several studies [4–8] have reviewed the advancements in this field (Table 1). Most studies focus on the application of machine learning (ML), deep learning (DL), and transfer learning (TL) techniques in signal processing (SP), feature extraction (FE), and the classification of AS or PS in the ICBHI2017 database.

Table 1. Review studies on RSA techniques.

	Involved Datasets	**Techniques**	**Purpose**	**Covered Years**
[4]	ICBHI2017, JUST Database, HF_Lung_V1/V2, and RespiratoryDatabase@TR	DL	AS and PS	2013–2023
[5]	ICBHI2017	FE and DL	AS and PS	2013–2023
[6]	ICBHI2017, HF_Lung_V1, R.A.L.E., and RespiratoryDatabase@TR	SP, FE, and DL	AS and PS	2011–2023
[7]	ICBHI2017 R.A.L.E., CheXpert, and ChestX-ray14	ML and DL	AS and PS	2015–2022
[8]	ICBHI2017, COUGHVID, Corp, and Coswara	FE, DL, and TL	AS and PS	2017–2022
ours	ICBHI2017	SP, FE, ML, DL, and TL	AS and PS	2017–2025

However, RSA techniques have rapidly evolved alongside artificial intelligence (AI) advancements, with notable progress in DL in recent years. To address this gap, the current study focuses on the ICBHI2017 database and examines SP, FE, ML, DL, and TL techniques with the goal of providing a comprehensive review on AS and PS classification, expanding the understanding of signal analysis techniques in related tasks.

1.2. The ICBHI2017 Database

Open-source databases are valuable for algorithm development, performance evaluations, and disease understanding. Table 2 presents details of open-source respiratory sound databases containing 1000 or more acoustic segments. It shows the number of acoustic segments (N_{seg}), the classification problems (C_{type}), the maximum number of categories (N_C), and additional information (S_{add}) regarding patient cases.

Table 2. Overview of open-sourced respiratory sound databases.

Database	N_{seg}	C_{type} (N_C)	S_{add}
RespiratoryDatabase@TR [9]	3696	AS (4)	Chest X-ray, heart sound, and questionnaire data
SPRSound [10]	9089	AS (6)	Demographics
HF_Lung_V1 [11]	9765	AS (4)	-
HF_Tracheal_V1 [12]	10,448	AS (3)	-
HF_Lung_V2 [13]	13,957	AS (5)	Demographics
ICBHI2017 [14]	6898	AS (4) and PS (7)	Demographics

A comprehensive review of RSA techniques using the ICBHI2017 database is timely and valuable for researchers and engineers who are interested in advancing this field. The ICBHI2017 database contains 6898 segments from 126 subjects, ranging from children to elderly individuals. The recordings were captured using various auscultation devices at multiple sites, with durations from 10 to 90 s. The acoustic segments are annotated and cross-reviewed by multiple experts. Demographic features are also provided for model training, subgroup analysis, and clinical modeling. In contrast, the HF_Lung_V2 database [13] consists of samples affected by issues such as missing annotations and subjective discrepancies. The SPRSound database [10] is limited to pediatric populations, and the HF_Tracheal_V1 database [12] involves anesthetized patients, both of which impede the generalization potential. In terms of classification, most studies focus on AS categorization [9–13], and only attention from [14] is additionally paid to PS recognition.

1.3. Literature Retrieval and Review of the ICBHI2017 Database

The keyword "ICBHI 2017" was searched using Google Scholar (accessed on 5 May 2025), and the database [14] has been used 517 times. Excluding non-English papers, books, editorials, review articles, dissertations, and other irrelevant literature, 135 technique papers remained. Figure 1 shows the RSA techniques reviewed in this study. The topics include signal processing, feature extraction, and sound classification, and detailed techniques are investigated and summarized under each topic.

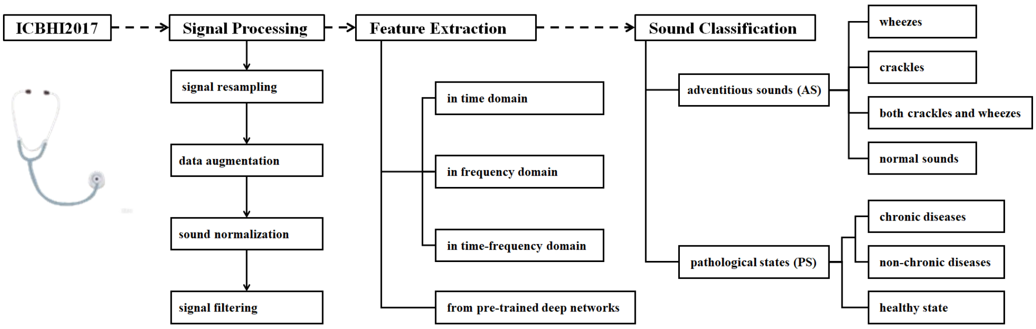

Figure 1. Review of RSA techniques based on the ICBHI2017 database.

2. Signal Processing

To ensure consistency, the data should undergo resampling, normalization, and filtering, with the purpose to prepare each segment for subsequent processing. Because of the class imbalance, data augmentation is introduced.

2.1. Signal Resampling

Signal resampling is applied to standardize the sampling rate, ensuring consistency across diverse data sources. Lower sampling rates improve computational and storage efficiency, whereas higher sampling rates preserve more detailed signal characteristics. Table 3 presents the resampling frequencies used in the literature, along with the corresponding references and the total number (#) of studies.

Table 3. Signal resampling frequencies.

Resampling Frequency	References (#)
1000 Hz	[15] (1)
2000 Hz	[5,16] (2)
4000 Hz	[17–54] (38)
6000 Hz	[55,56] (2)
8000 Hz	[57–64] (8)
10,000 Hz	[19,40] (2)
16,000 Hz	[28,65–71] (8)
22,050 Hz	[72–78] (7)
32,000 Hz	[79] (1)
44,100 Hz	[6,19,40,78,80–83] (8)

The resampling frequencies range from 1000 Hz to 44,100 Hz, and 4000 Hz is the most widely used in 38 studies. Different sampling rates are chosen for varying task requirements and signal characteristics. Selecting an appropriate sampling rate requires balancing signal fidelity, computational efficiency, and model training performance.

2.2. Data Augmentation

The class imbalance is present in the ICBHI2017 database, and 5641 out of 6898 respiratory cycles (81.8%) belong to the COPD category, making data augmentation essential for reducing the risk of overfitting and improving the model robustness [84]. To address this issue, standard data augmentation techniques are used.

Time stretching involves stretching or compressing the duration of the signal without changing its pitch. It benefits feature extraction by generating multi-scale samples of the original signal. The operation of time stretching is formulated as shown in Equation (1),

$$x(t) = x_0(at), \tag{1}$$

where a is the time-scaling coefficient. When $a < 1$, the signal duration increases and playback slows; when $a > 1$, the duration decreases and playback speeds up; when $a = 1$, the signal is unchanged.

Pitch shifting changes the pitch of the input signal without altering its duration. It enables the generation of signals with different pitches. The operation of time stretching can be formulated as shown in Equation (2),

$$x(t) = x_0(t) \cdot e^{j2\pi f_0 t}, \tag{2}$$

where f_0 denotes the pitch offset.

Adding noise involves inducing noise into the training samples. It helps the simulation of background interference in real-world environments or under noisy conditions [85]. The operation of adding noise can be formulated as shown in Equation (3),

$$x(t) = x_0(t) + \eta(t), \tag{3}$$

where $\eta(t)$ stands for adding noise.

Speed transformation simulates different breathing frequencies and rhythms by changing the playback speed of the signals with varying factors. It helps reduce the risk of overfitting [86]. The speed transformation operation can be formulated as shown in Equation (4),

$$x(t) = x_0\left(\frac{t}{b}\right), \tag{4}$$

where b is the speed coefficient. When $b > 1$, playback speeds up and the duration shortens; when $b < 1$, playback slows and the duration lengthens; and when $b = 1$, the signal remains unchanged.

Time shifting involves shifting the signal along the time axis either forward or backward. It simulates different starting points or variations in breathing. This technique increases the diversity of the signals to help a model better adapt to different breathing rhythms and temporal changes [86]. Equation (5) shows the operation of time shifting,

$$x(t) = x_0(t + \Delta t), \tag{5}$$

where Δt denotes a time shift.

Dynamic range compression reduces the parts of the signal with large volume variations. It makes the overall volume more balanced with a reduced impact on background noise, thereby highlighting the details of the respiratory sounds [87]. The operation of dynamic range compression can be formulated as shown in Equation (6),

$$x(t) = \text{sign}(x(t)) \cdot \log(1 + |x(t)|), \tag{6}$$

where $\text{sign}(x(t))$ represents the signal symbol.

Frequency masking works by masking a given frequency interval. It operates by masking a specified time interval to simulate signal loss or noise interference that may occur in real-world scenarios [88].

Two masking techniques help the model handle real-world challenges such as noise, signal loss, or missing signals. The operation of frequency masking in the frequency domain can be formulated as shown in Equation (7),

$$S_{\text{masked}}(f,t) = \begin{cases} 0 & \text{if } f_0 \leq f \leq f_0 + \Delta f \\ S(f,t) & \text{otherwise,} \end{cases} \tag{7}$$

where f_0 is the starting frequency index of the mask, and Δf is the width of the mask (the number of frequency channels).

The other operation of frequency masking in the time domain is shown in Equation (8),

$$S_{\text{masked}}(f,t) = \begin{cases} 0 & \text{if } t_0 \leq t \leq t_0 + \Delta t \\ S(f,t) & \text{otherwise,} \end{cases} \tag{8}$$

where t_0 is the starting time index of the mask, and Δt is the width of the mask or the number of time steps.

Table 4 summarizes the applications of standard data augmentation techniques used on the ICBHI2017 database. It is found that time stretching, adding noise, and pitch shifting are widely applied, followed by time shifting and speed transformation.

Table 4. Standard data augmentation techniques.

Standard Techniques	**References (#)**
Time stretching	[4,40,41,62–65,72,73,76,78,89–97] (20)
Pitch shifting	[4,26,38–41,62,63,72,83,93,96–100] (16)
Adding noise	[26,32,38–41,62–64,76,82,89–93,99,100] (18)
Speed transformation	[26,38,39,75,99] (5)
Time shifting	[32,39,63,89,92,100–102] (8)
Dynamic range compression	[4,72,97] (3)
Frequency and time masking	[27,70,95] (3)

Time stretching, adding noise, and pitch shifting are widely used as standard data augmentation techniques for alleviating the class imbalance issue. Several other techniques have also been proven effective. The variance time–length product method applies a random envelope factor to each recording and maps signal frequencies to new ones [103]. This technique has been used in [38,73,92,94,104,105]. The mix-up method enhances data diversity by linearly interpolating between two random samples to generate a new one [106]; it has been broadly adopted in respiratory sound processing [28,62,66]. In addition, strategies such as intelligent padding and random sampling [26], slicing and feature fusion [107], patch random selection with positional encoding [70], and the use of the Griffin–Lim algorithm [108] have demonstrated benefits for augmenting respiratory sound data.

2.3. Signal Normalization

Signal normalization includes both duration normalization and amplitude normalization. The former aligns signals of varying lengths to a consistent time scale, while the latter scales the signal amplitudes to a uniform range. Both steps aim to ensure the comparability of respiratory signals across different samples.

2.3.1. Duration Normalization

Duration normalization involves adjusting the length of respiratory sound recordings to a fixed duration. This process eliminates the variability caused by signal lengths and ensures that all samples can be compared on the same temporal scale. Common standardization methods include cropping and padding [15,19,20,23,37,40,44,48–51,59,62,66,70,74,77,80,95,99–101,109–113]. The cropping method is used to truncate signals that exceed the target duration, while padding adds silent sections (typically zeros) to shorter signals to extend them to the desired length. These procedures enable a consistent comparison and analysis within a unified time window, preserving the temporal structure needed for downstream processing and model training.

2.3.2. Amplitude Normalization

The purpose of amplitude normalization is to map the signals' amplitudes to a uniform range to eliminate variability caused by amplitudes across signals. One widely used normalization method involves scaling the amplitude of the signal proportionally to a specified range [5,15,27,29,35,44,45,53,58,59,62,64,68,72,78,94,96,110,114].

Another method is root mean square (RMS) normalization [97], which adjusts the signal's RMS value to standardize the amplitude to enhance the comparability of the signals across different environments. Equation (9) formulates the operation,

$$\mathrm{RMS}(x) = \sqrt{\frac{1}{d}\sum_{i=1}^{d} x_i^2}, \tag{9a}$$

$$\mathrm{RMS}_{\mathrm{acum}}(x) = \frac{x}{\mathrm{RMS}(x)} \odot \gamma, \tag{9b}$$

in which x is the input vector, d is the dimension of the vector, and γ is a trainable parameter vector initialized randomly and optimized during the training process. The operator $\odot$ denotes element-wise multiplication.

Min-max normalization [50,115–117] maps the signal values to [0, 1]. It preserves the relative proportionality of the data through linear transformation. This technique enhances data consistency and maintains the relative structure of signal features. Equation (10) shows the normalization procedure,

$$x = \frac{x_t - \min(x_t)}{\max(x_t) - \min(x_t)}, \tag{10}$$

in which $\min(x_t)$ and $\max(x_t)$ correspond to the minimum value and the maximum value of all samples in the dataset.

In addition, z-score normalization [31,55] converts the signals into a form with zero mean and unit standard deviation. It reduces the differences in the scale and improves data consistency and model performance. Equation (11) shows the operation of z-score normalization,

$$z = \frac{x - \mu(x)}{\sigma(x)}, \tag{11}$$

where $\mu(x)$ and $\sigma(x)$ correspond to the mean and standard deviation of the signal x.

2.4. Signal Filtering

Due to different environmental conditions, the raw respiratory signals may contain noise and irrelevant information, and signal filtering or denoising becomes crucial with the purpose to filter out noise and irrelevant components, and therefore, features related to respiratory activity can be highlighted to provide reliable subsequent analysis.

2.4.1. Environmental Noise Suppression

Environmental noise suppression uses filtering techniques to eliminate or reduce external environmental noise interference to extract a clear target signals. It implements high-pass filters [69,95], which can effectively remove low-frequency noise. Typical high-pass filter designs include finite impulse response filters [18,101,110], Butterworth filters [53,80], and Bessel filters [65]. These filters preserve signal waveform characteristics and provide proper phase responses for signal processing outcomes.

2.4.2. Heart Sound Interference Removal

Heart sound interference removal aims to eliminate heart sound components that overlap with respiratory sounds. Common methods include the use of band-pass filters [15,19,21,22,30,58]. They retain the target signal by selecting specific frequency bands and suppressing heart sound interference. Among band-pass filters, the Butterworth filter is one of the most used designs [17,23,26,34,38,42,44,45,49,52,55–57,60,64,114,117,118], and smooth frequency response characteristics allow for the effective attenuation of unwanted frequency bands [119]. Additionally, anti-aliasing low-pass filters [120] can remove high-frequency noise and ensure minimal impact on the target signal. Another common signal smoothing method is the Savitzky–Golay filter [121]. It improves the signal-to-noise ratio by removing noise without compromising the overall morphology of the signals [122].

3. Feature Extraction of Respiratory Signals

Feature extraction is a core step for identifying different types of respiratory signals. The techniques could be categorized into feature extraction in domains, including the time domain, the frequency domain and the time–frequency domain; feature extraction from nonlinear time series; and feature extraction using pre-trained deep learning networks. These methods capture the multi-dimensional cues of respiratory signals and provide important features for subsequent classification tasks by using machine learning-based or hybrid learning-based approaches.

3.1. Feature Extraction in Domains

3.1.1. Feature Extraction in the Time Domain

In the time domain, features are extracted to embed the temporal characteristics and dynamic changes of signals. Statistical features, such as the mean, variance, maximum, and minimum values, provide the energy levels and the range of variations to understand the overall behavior of the signals [123].

Shannon entropy acts as a measure of signal uncertainty, randomness, or complexity that reflects the distribution characteristics of information within the signal. It is an effective, quantitative basis for signal analysis [123]. Equation (12) shows the computing of Shannon entropy,

$$H(x) = -\sum_{i=1}^{N} p(x_i)\log(p(x_i)), \tag{12}$$

where $H(x)$ denotes the entropy of the random variable x, and $p(x_i)$ presents the probability of the occurrence of the i-th event.

The zero-crossing rate (ZCR) refers to the number of times a signal crosses the zero axis. It provides a quantitative measure of the signal's periodicity. It is helpful in analyzing the periodicity and noise characteristics of audio signals. It is formulated as shown in Equation (13),

$$ZCR = \frac{1}{2(N-1)}\sum_{i=1}^{N-1} |\mathrm{sign}(x_i) - \mathrm{sign}(x_{i+1})|, \tag{13}$$

where x_i is the signal value at the i-th sample point, and sign() represents the sign function.

The methods for extracting the time-domain features of respiratory signals in the ICBHI2017 database are shown in Table 5. It is found that statistical features are preferred in time-domain-based feature extraction, and ZCR and Shannon entropy are also used.

Table 5. Feature extraction of respiratory signals in the time domain.

Time-Domain-Based Features	References (#)
Statistical features	[4,22,29,46,47,57,58,81,90,100,108,124–126] (14)
Shannon entropy	[82,127] (2)
Zero-crossing rate (ZCR)	[4,29,56,82,100,124] (5)

3.1.2. Feature Extraction in the Frequency Domain

In the frequency domain, the frequency distribution and energy characteristics of signals are explored. As one of the core methods, the Fourier transform (FT) decomposes the time-domain signal into sine wave components of different frequencies, and the spectral representations of the signal are obtained. The discrete FT is shown in Equation (14),

$$X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x[n] \cdot e^{-j\omega n}, \tag{14}$$

where ω denotes the angular frequency, and $x[n]$ is the value of the n-th sampling point.

Spectral features, such as spectral centroid, spectral bandwidth, and spectral flatness, provide a straightforward reflection of the frequency structure of the signal by analyzing the energy distribution statistically in the spectrum. Due to their simple computation and ease of implementation, spectral features act as a benchmark for an initial description and comparison of the frequency-domain characteristics of the signals [128].

Power spectral density analysis is used to describe the energy distribution of signals across different frequencies. By performing FT or auto-correlation analysis, the power spectral density can be estimated to quantify the characteristics of the power distribution to the frequency components.

In the ICBHI2017 database, the methods for frequency-domain feature extraction are shown in Table 6. The spectral features are preferred over others, followed by FT features and power spectral density analysis.

Table 6. Feature extraction of respiratory signals in the frequency domain.

Frequency-Domain-Based Features	References (#)
Fourier Transform (FT)	[65,90] (2)
Spectral features	[4,21,29,81,83,108,129] (7)
Power spectral density analysis	[55,124,130] (3)

3.1.3. Feature Extraction in the Time–Frequency Domain

Feature extraction in the time–frequency domain explores how to perceive the temporal variations and frequency distribution characteristics of signals simultaneously. Mel-frequency cepstral coefficients (MFCCs) mimic the auditory characteristics and convert the spectrum into cepstral coefficients on the Mel scale to capture the timbral features. The computing of MFCCs can be described as shown in Equation (15),

$$MFCC_n = \sum_{k=1}^{K} \log|X_k| \cdot \cos\left(n \cdot \frac{(k-1)\pi}{K}\right), \tag{15}$$

where K represents the number of Mel filter bank bands, X_k denotes the energy of the k-th Mel frequency band, and $\cos(\cdot)$ refers to the discrete cosine transform.

The short-time Fourier transform (STFT) slides a window function along the time axis, performs the Fourier transform on the signal within each window, and obtains the time–frequency representation of the signal. It can be formulated as shown in Equation (16),

$$X(t,f) = \sum_{n=-\infty}^{\infty} x[n] \cdot w[n-t] \cdot e^{-j2\pi fn}, \tag{16}$$

where $x[n]$ is the original signal, and $w[n-t]$ is the window function.

A spectrogram is a visual tool used to present the frequency distribution of a signal intuitively. After applying the FT on the signal, it arranges the spectrum of each moment in chronological order, and the energy intensity of each frequency component is visualized by using different colors or grayscale levels. Subsequently, the frequency structure of the signal and its dynamic changes over time are uncovered [129].

The Mel spectrogram further maps the frequency axis to the Mel scale, and nonlinear frequency perception can be better simulated [131]. The logarithmic Mel spectrogram extracts the spectrum of the signal through the STFT, adjusts the spectrum using Mel filters to align with human auditory perception, and enhances the distinguishability of weak signals through logarithmic transformation.

A wavelet transform (WT) performs multi-scale signal analysis through the dilation and translation of wavelet functions. It can offer high time and frequency resolution for non-stationary signal analysis [132]. The operation of the WT can be formulated as shown in Equation (17),

$$W_{\Phi}(s,\tau) = \int_{-\infty}^{\infty} x(t)\cdot \Phi^{*}\left(\frac{t-\tau}{s}\right)dt, \tag{17}$$

where $x(t)$ is the original signal, $\phi(t)$ denotes the mother wavelet function, s is the scale parameter, and τ is the translation parameter.

The gammatonegram utilizes a gammatone filter bank to decompose the signal to extract the energy distribution of each frequency channel, and it can visually present the time–frequency dynamics of the signal.

The constant-Q transform uses a logarithmic scale for frequency resolution to provide higher frequency resolution at low frequencies and better time resolution at high frequencies. It is suitable for the analysis of complex frequency components in respiratory signals [133].

As shown in Table 7, based on the ICBHI2017 database, the respiratory signal analysis in the joint time–frequency domain includes various types of signal transformation for informative feature extraction. The STFT is preferred among the methods, followed by Mel spectrograms, MFCCs, and WT-based features.

Table 7. Feature extraction of respiratory signals in the joint time–frequency domain.

Time–Frequency-Domain-Based Features	References (#)
Mel-Frequency Cepstral Coefficients (MFCCs)	[4,20,33,52,62,64,76,104,110,121,134–136] (13)
Mel Spectrograms	[4,6,17,18,22,29,41,46,56,58,63,64,74,82,83] (15)
Logarithmic Mel Spectrograms	[34,38,51,63,67,68,71] (7)
Short-Time Fourier Transform (STFT)	[4,6,16,19–21,23–25,33,35,37,40,44,45,53,63,64,69,76,78,83,97,104,105,110,112,117,118,137] (30)
Wavelet Transform (WT)	[4,5,20,25,28,32,40,53,60,80,81,114,138] (13)
Wavelet Packet Integral	[6,30,58,111,135] (5)
Constant-Q Transform	[25,41,92,99,139] (5)
Gammatonegram	[27,28,66,92,107] (5)
Mel Filter Bank	[28,55,61,65,69,105] (6)

3.2. Feature Extraction from Pre-Trained Deep Neural Networks

Deep convolutional neural networks (CNNs) have demonstrated powerful capacities in hierarchical abstract feature representation in a broad range of applications, and a number of pre-trained deep networks have been used in the feature extraction of respiratory signals. Through specific convolutional kernels, SincNet extracts features for respiratory signal representation [77]. Fraiwan et al. combine both 1D CNN and bi-directional long short-term memory (Bi-LSTM) models for temporal modeling in which the 1D CNN is able to capture the local description, while Bi-LSTM is used to obtain the bi-directional temporal dependencies [31]. Meanwhile, EasyNet [50] and parallel autoencoders [40] extract low-dimensional features through unsupervised learning to uncover the latent structure of respiratory signals. Self-attention mechanisms used in densely connected networks [140], the audio spectrogram transformer (AST) [141], and novel networks emphasize key information that excels at capturing long-range dependencies [105]. In addition, the combination of conditional Gaussian capsule networks with cubic encoders enhances respiratory signal representation learning, since dynamic routing and multi-dimensional mapping are embedded [54].

4. Learning-Based Respiratory Sound Classification

Despite signal processing and feature extraction, respiratory signal classification methods have evolved from machine learning and deep learning to hybrid learning, and the training strategies involve supervised learning, self-supervised learning, contrastive learning, and transfer learning.

4.1. Performance Evaluation Metrics

For four-class prediction, assuming that P_n, P_c, P_v, and P_b, respectively, denote the number of correctly predicted samples for the classes of "normal," "crackle," "wheeze," and "both" (i.e., samples exhibiting both crackle and wheeze), N_n, N_c, N_v, and N_b stand for the number of samples in each respective class.

In the challenge based on the ICBHI2017 database, the metric specificity (SPE) measures the capacity of a model to correctly identify healthy samples, and it is formulated as shown in Equation (18),

$$SPE = \frac{P_n}{N_n}. \tag{18}$$

The second metric, sensitivity (SEN), measures the model's ability to correctly identify pathological samples. It is computed as shown in Equation (19),

$$SEN = \frac{P_c + P_v + P_b}{N_c + N_v + N_b}. \tag{19}$$

Accuracy (ACC) estimates the overall classification correctness of the model, and its formula is shown in Equation (20),

$$ACC = \frac{P_n + P_c + P_v + P_b}{N_n + N_c + N_v + N_b}. \tag{20}$$

The challenge also provides an ICBHI score (HS) that considers both specificity and sensitivity to evaluate the overall performance as shown in Equation (21),

$$HS = \frac{1}{2} \times (SPE + SEN). \tag{21}$$

These metrics are widely used for performance evaluations and comparisons that can be computed in a similar way for binary classification, ternary classification, and multi-class classification tasks [142,143].

4.2. Machine Learning-Based Respiratory Sound Classification

After features are handcrafted, machine learning-based respiratory signal classification typically relies on the selection and training of machine learning classifiers. Constrained by domain knowledge and classifier exploration, the performance remains unsatisfactory, and several works are shown in Table 8, where # stands for the number of classes. Widely used ML classifiers include a support vector machine (SVM), the hidden Markov model

(HMM), the Gaussian mixture module (GMM), and tree models (such as the RUSBoost tree and random forest), and *k*-fold cross-validation (*k*-FCV) is widely used for random data splitting.

Table 8. Machine learning-based respiratory sound classification.

		Year	#	Splitting	SEN (%)	SPE (%)	ACC (%)	HS (%)
[58]	RUSBoost Tree	2019	2	-	93.60	86.80	87.10	90.20
[17]	SVM	2018	4	-	77.80	48.90	49.98	-
[18]	HMM/GMM	2018	4	10-FCV	-	-	-	39.56
[22]	Random Forest	2020	7	70-30-0	-	-	88.00	87.00

4.3. Deep Learning-Based Respiratory Sound Classification

Deep learning has updated the performance in massive applications. In the field of respiratory signal classification, the CNN, the RNN (recurrent neural network), and their variants have become the mainstream methods that learn feature representation and signal classification in an end-to-end manner.

A CNN learns representative features through convolutional and pooling layers. It is particularly suitable for processing spectrograms or time–frequency representations of respiratory sounds. Table 9 shows the performance of CNN-related models for respiratory signal classification. Except for novel designs of CNN architectures and different splitting ratios, ResNet [144] and VGG [145] are the most widely applied models in binary, ternary, and multi-class prediction [23,26,36,45,93,94,118]. Meanwhile, promising results have been obtained in binary classification ([93] with ACC and HS both obtaining $\geq$ 95.00%) and ternary classification ([115] with all metrics obtaining $\geq$ 98.00%), while there is room for further improvement in four-class prediction. (The highest HS value is $\leq$ 80.00%.) Even though with the use of the attention mechanism, the studies remain insufficient for six-class and eight-class predictions.

Table 9. CNN-based respiratory sound classification.

		Year	#	Splitting	SEN (%)	SPE (%)	ACC (%)	HS (%)
[134]	CNN	2018	2	80-20-0	-	-	78.00	85.00
[93]	ResNet-34	2022	2	80-20-0	-	-	96.00	95.80
[134]	CNN	2018	3	80-20-0	-	-	82.00	84.00
[115]	CNN	2020	3	10-FCV	98.60	98.80	99.40	98.70
[28]	CNN	2021	3	60-40-0	88.00	85.00	86.00	86.50
[60]	CNN	2021	3	70-10-20	-	-	98.70	98.47
[102]	CNN	2022	3	70-20-10	-	-	-	99.30
[118]	Bi-ResNet	2019	4	60-40-0	31.12	69.20	57.29	50.16
[73]	CNN	2020	4	80-20-0	87.30	69.40	-	78.35
[23]	ResNet-18	2020	4	70-30-0	81.25	17.84	-	49.55
[26]	ResNet-34	2021	4	60-40-0	39.00	71.40	-	55.20
[26]	ResNet-34	2021	4	80-20-0	78.80	53.60	-	66.20
[28]	CNN	2021	4	60-40-0	32.00	73.00	-	53.00
[94]	ResNet-50	2022	4	60-40-0	37.24	79.34	-	58.29
[34]	CNN	2022	4	60-40-0	27.78	72.96	-	50.37
[36]	ResNet	2022	4	60-40-0	30.00	70.00	-	50.00
[45]	ResNet-34	2023	4	60-40-0	25.10	75.30	-	50.20
[45]	ResNet-34	2023	4	80-20-0	79.56	57.89	-	68.72
[72]	2D-CNN	2019	6	70-30-0	-	-	97.00	-
[146]	CNN	2024	8	80-20-0	99.42	96.53	96.03	97.99
[78]	VGG16	2024	8	60-40-0	-	-	75.00	72.00

The RNN and its variants, such as the gated recurrent unit (GRU), are well-suited for processing sequential signals and temporal characteristics. Table 10 shows the results when using the RNN and its variants on the database. Notably, LSTM and Bi-LSTM are widely used, and high ACC values are obtained for binary, six-class, and eight-class prediction tasks.

Table 10. Respiratory sound classification using the RNN and its variants.

		Year	#	Splitting	SEN (%)	SPE (%)	ACC (%)	HS (%)
[109]	LSTM	2019	2	80-20-0	82.00	99.00	99.00	91.50
[101]	RNN	2018	4	5-FCV	74.10	61.70	67.90	67.90
[108]	Bi-GRU	2020	4	70-30-0	80.65	64.24	64.94	72.45
[136]	BiLSTM-BiGRU	2021	6	75-12.5-12.5	-	-	96.20	-
[107]	Bi-LSTM	2021	6	80-20-0	-	-	88.30	-
[40]	LSTM	2024	8	80-20-0	99.10	89.56	94.16	94.33

4.4. Hybrid Learning-Based Respiratory Sound Classification

The hybrid models combine the feature extraction capability of the CNN and the temporal modeling ability of the RNN to enhance the classification performance. Table 11 shows these hybrid models in respiratory signal classification. The results are promising on binary classification (ACC $\geq$ 94.00%), on ternary classification (ACC $\geq$ 90.00%), on six-class prediction (metrics $\geq$ 96.00%), and on eight-class prediction (ACC $\geq$ 86.00%).

Table 11. Hybrid learning-based respiratory sound classification.

		Year	#	Splitting	SEN (%)	SPE (%)	ACC (%)	HS (%)
[69]	VGGish-BiGRU	2023	2	85-15-0	-	-	94.00	94.00
[115]	CNN-LSTM	2022	3	-	99.15	97.60	99.62	-
[33]	CNN-LSTM	2022	4	10-FCV	84.26	52.78	76.39	68.52
[31]	CNN-BiLSTM	2022	6	10-FCV	99.69	98.43	99.62	99.06
[89]	CNN-LSTM	2022	8	-	97.71	84.73	82.35	-
[89]	CNN-RNN	2022	8	-	-	-	86.00	87.00

4.5. Transformer-Based Respiratory Sound Classification

The transformer leverages self-attention to capture the global temporal features of respiratory rhythms. By incorporating positional encoding, it can preserve respiratory phase information that enhances the identification of long-range dependencies and improves the classification accuracy of abnormal breath sounds. The transformer-based respiratory signal classification based on the database is shown in Table 12. The vision transformer (ViT) [147] and AST are preferred on four-class prediction, and there is sufficient room for further improvement in the classification performance (HS $\leq$ 70.00%).

Table 12. Transformer-based respiratory sound classification.

		Year	#	Splitting	SEN (%)	SPE (%)	ACC (%)	HS (%)
[148]	ViT	2022	4	60-40-0	36.41	78.31	-	57.36
[112]	ViT	2023	4	80-20-0	-	-	-	69.30
[46]	AST	2023	4	60-40-0	42.91	62.11	53.53	50.76
[71]	AST	2024	4	60-40-0	44.37	80.43	-	62.40

5. Current Achievement on the Respiratory Sound Classification

The ultimate goal of respiratory signal processing, feature extraction, machine learning, and deep learning is accurate classification. The ICBHI2017 database supports both AS categorization and PS recognition tasks. The former categorizes respiratory signals into “normal”, “crackle”, “wheeze”, and “both crackle and wheeze” groups. The latter is used to distinguish between "healthy" and "unhealthy" cases which can be further refined by subdividing the unhealthy category into more specific diagnostic groups.

5.1. Performance on AS Categorization

The database includes 3642 normal segments (52.8% samples), 1864 crackle segments, 886 wheeze segments, and 506 segments (7.3% samples) containing both crackles and wheezes. According to the literature, there are two groups of approaches for AS categorization based on data splitting. One group uses the official split, where the training and testing sets are predefined and fixed. The other group involves custom splits, where researchers design their own data partitions with varying ratios for training, validation, and testing.

5.1.1. Performance on AS Classification When Using the Official Data Split

Table 13 shows the classification performance, and the database is split into a training set (60% samples) and a testing set (40% samples) with officially fixed cases. The highest metric values is highlighted in bold.

Table 13. Performance of AS categorization using official data splitting.

	Backbone	Year	Splitting	SEN (%)	SPE (%)	HS (%)
[118]	Bi-ResNet	2019	60-40	31.12	69.20	50.16
[117]	LungRN+NL	2020	60-40	41.32	63.20	52.26
[55]	CNN	2021	60-40	42.00	42.00	42.00
[26]	ResNet-34	2021	60-40	39.00	71.40	55.20
[66]	CNN-MoE	2021	60-40	26.00	68.00	47.00
[28]	CNN-Inception	2021	60-40	32.00	73.00	53.00
[74]	ARSC-Net	2021	60-40	46.38	67.13	56.76
[148]	ViT	2022	60-40	36.41	78.31	57.36
[94]	ResNet-50	2022	60-40	37.24	79.34	58.29
[34]	CNN	2022	60-40	27.78	72.96	50.37
[36]	ResNet	2022	60-40	30.00	70.00	50.00
[62]	ResNeStIBN	2022	60-40	40.20	70.40	55.30
[149]	CNN	2022	60-40	39.15	76.93	58.04
[99]	GTFA-Net	2023	60-40	**48.40**	72.10	60.25
[45]	ResNet-34	2023	60-40	25.10	75.30	50.20
[46]	FNN	2023	60-40	44.55	55.95	50.25
[114]	BLNet	2023	60-40	42.63	61.33	51.98
[150]	AST	2023	60-40	43.07	81.66	62.37
[52]	OFGST-Swin	2024	60-40	40.53	71.56	56.05
[151]	CLAP	2024	60-40	45.67	81.40	**63.54**
[71]	AST	2024	60-40	44.37	80.43	62.40
[92]	CycleGuardian	2025	60-40	44.47	**82.06**	63.26

The highest SEN, SPE, and HS values are 48.40%, 82.06%, and 63.54%, respectively, achieved by GTFA-Net [99], CycleGuardian [92], and CLAP [151]. Five models obtain HS $\geq$ 60.00% among the algorithms. Technically, GTFA-Net [99] develops group-wise time–frequency attention that segments Mel spectrograms into different groups by frequency-dimension random masking, and then, the states of the groups are extracted, weighted, and aggregated into global representations for AS classification. CycleGuardian [92] integrates multi-channel spectrograms, adopts time-grouped encoding, and combines deep clustering with group-mixed contrastive learning, and group feature embedding and cluster-projection fusion are incorporated into a multi-objective optimization manner for improved performance and generalization. CLAP [151] designs contrastive learning and multi-modal fusion, and a pre-trained framework is employed to align respiratory sounds and textual metadata in a shared feature space. It can handle missing or unseen metadata, encode key variables, and mitigate variability caused by different devices and recording positions.

5.1.2. Performance on AS Categorization When Using Custom Data Splits

Except for officially fixed data splitting as used in the challenge, numerous algorithms use different kinds of splitting strategies, including *k*-FCV and different splitting ratios for training, validating, and testing. The performance of AS categorization using wild data splitting is summarized in Table 14.

Table 14 shows that the ResNet-based model [36] obtains the highest SEN (93.00%) and HS (88.00%) using 10-FCV, the SincNet-based model [77] achieves the best SPE (95.00%) and ACC (91.13%) when using 80% samples for model training, and several models [36,77,80] lead to HS values larger than 80.00%, which is much higher than those top-ranking models that use the official data splitting strategy (Table 13). Notably, the ResNet-based model [36] combines fluid–structure interaction dynamics to simulate the coupled bronchial airflow and wall deformation and to enhance sound source modeling accuracy, and ResNet is integrated to incorporate channel-wise and frequency-band attention for multi-dimensional feature enhancement and final classification. The SincNet-based model [77] is a two-stage self-supervised contrastive learning framework. The first stage involves a waveform encoder to extract informative frequency components, and the encoder is pre-trained on a large-scale dataset to learn robust and generalizable audio representations. The second stage introduces a contrastive variational autoencoder that leverages latent variable modeling to address the class imbalance.

Table 14. Performance of AS categorization using custom data splits.

	Backbone	**Year**	**Splitting**	**SEN (%)**	**SPE (%)**	**ACC (%)**	**HS (%)**
[101]	RNN	2018	5-FCV	74.10	61.70	67.90	67.90
[123]	CNN-RNN	2020	80-20-0	84.14	48.63	58.47	66.38
[19]	CNN	2020	10-FCV	86.00	61.00	69.00	65.00
[95]	InfoGAN	2020	5-FCV	70.20	79.30	-	74.80
[117]	LungRN+NL	2020	5-FCV	63.20	41.32	-	52.26
[20]	VGG-16	2020	5-FCV	66.80	47.40	55.60	57.10
[73]	CNN	2020	80-20-0	87.30	69.40	-	78.35
[23]	ResNet18	2020	70-30-0	81.25	17.84	-	49.55
[108]	BiGRU-XGBoost	2020	70-30-0	80.65	64.24	64.94	72.45
[25]	CRNN	2021	5-FCV	83.00	64.00	-	74.00
[26]	ResNet-34	2021	80-20-0	78.80	53.60	-	66.20
[66]	CNN-MoE	2021	5-FCV	86.60	71.30	-	78.90
[67]	CNN	2021	80-20-0	85.44	70.93	78.73	78.18
[80]	CNN-RNN	2021	5-FCV	90.66	72.32	-	81.64
[74]	ASRC-Net	2021	5-FCV	74.76	58.95	-	66.86
[33]	CNN-LSTM	2022	10-FCV	84.26	52.78	76.39	68.52
[36]	ResNet	2022	5-FCV	87.00	80.00	-	83.00
[36]	ResNet	2022	10-FCV	**93.00**	84.00	-	**88.00**
[61]	MBTCNSE	2023	80-20-0	86.10	65.30	72.50	75.70
[99]	GTFA-Net	2023	80-20-0	82.00	61.40	-	71.70
[77]	SincNet	2023	80-20-0	80.20	**95.00**	**91.30**	87.60
[77]	SincNet	2023	10-FCV	77.60	94.90	90.60	86.30
[37]	CNN	2023	10-FCV	68.84	52.71	62.93	60.78
[45]	ResNet34	2023	80-20-0	79.56	57.89	-	68.72
[114]	BLNet	2023	80-20-0	79.13	66.31	-	72.72
[40]	LSTM	2024	80-20-0	92.49	89.56	79.61	78.67

5.2. Performance on PS Recognition

According to the number of pathological conditions considered, PS recognition can be categorized into binary, ternary, and multi-class classification. Binary classification typically distinguishes between healthy and unhealthy states, providing a straightforward diagnostic decision. Ternary classification further divides the unhealthy category into two distinct pathological groups, enabling more refined differentiation of disease severity or type.

Multi-class classification expands this approach by identifying multiple specific conditions, offering detailed diagnostic insights that can support targeted treatment planning and personalized healthcare.

5.2.1. Performance on PS Binary Classification

As for binary classification, the ICBHI2017 dataset contains 35 healthy cases (3.8% of the samples) and 885 unhealthy cases (96.2% of the samples). PS binary classification is particularly suitable for the preliminary screening phase of diseases. Table 15 shows the performance of different methods for this task on the dataset.

Table 15. Performance of PS binary classification.

	Backbone	Year	Splitting	SEN (%)	SPE (%)	ACC (%)	HS (%)
[134]	CNN	2018	80-20-0	-	-	78.00	85.00
[58]	RUSBoost	2019	-	93.60	86.80	87.10	90.20
[109]	LSTM	2019	80-20-0	82.00	99.00	99.00	92.00
[66]	CNN-MoE	2021	5-FCV	86.00	98.00	-	92.00
[152]	SVM	2022	10-FCV	96.60	**100.0**	-	98.30
[93]	ResNet34	2022	80-20-0	-	-	96.00	95.80
[69]	VGGish-StackedBiGRU	2023	85-15-0	-	-	94.00	94.00
[48]	DNN	2024	70-30-0	68.00	**100.0**	96.00	84.00
[50]	EasyNet	2024	5-FCV	99.00	99.00	99.70	99.00
[137]	MHSONN	2024	80-10-10	**99.73**	99.85	**99.81**	**99.79**

High performance is achieved on the PS binary classification task. MHSONN [137] achieves the highest SEN, ACC, and HS values, as well as the second-best SPE value, all of which are larger than 99.00%. It integrates time–frequency representations from Mel spectrograms, constant-Q transform spectrograms, and Mel-frequency cepstral coefficients to capture both the frequency-domain dynamics and nonlinear characteristics; employs a self-organizing operational neural network via generative operational perceptrons; and utilizes a multi-head architecture to process multi-modal features in parallel for global state recolonization. A comparable model is the EasyNet model [50] that designs a streamlined hierarchical architecture with targeted parameters. In the architecture, the first stage captures fundamental frequency components, while the second stage utilizes depth-wise separable convolutions to extract temporal features from high-frequency components, and average pooling is used to compress the feature space for PS classification.

5.2.2. Performance on PS Ternary Classification

In PS ternary classification, the unhealthy cases are further divided into the cases with chronic diseases or non-chronic diseases, which differ significantly in clinical treatment and management strategies. Specifically, the database contains 35 healthy cases (3.8% of the samples), 75 non-chronic cases (8.2% of the samples), and 810 chronic cases (88.0% of the samples). Table 16 shows the performance of PS ternary classification.

The CNN-VAE-based model [115] obtains metric values larger than 98.00%, and the CNN-LSTM-based model [89] performs well. The CNN-VAE model [115] employs a variational autoencoder to generate synthetic data samples, and a Kullback–Leibler divergence regularization term is introduced to constrain the latent variables following a standard normal distribution. It enables the effective augmentation of minority class samples and improves the classification performance. The CNN-LSTM-based model [89] adopts a 1D CNN architecture for feature extraction with various activation functions, and the features are fed into LSTM to model the temporal dependencies, long-term dependencies, and dynamic variations within the signals for improved PS prediction. Notably, a built-in 11-layered network [102] achieves the highest HS value. It leverages auditory perception, frequency energy distribution, and pitch contour statistics to enhance feature expressiveness and introduces a delayed superposition augmentation method to enrich the data samples by overlapping time-shifted signals.

Table 16. Performance on PS ternary classification.

	Backbone	Year	Splitting	SEN (%)	SPE (%)	ACC (%)	HS (%)
[134]	CNN	2018	80-20-0	-	-	82.00	84.00
[109]	LSTM	2019	80-20-0	82.00	98.00	98.00	91.00
[115]	CNN-VAE	2020	10-FCV	98.60	**98.80**	**99.40**	98.70
[28]	CNN	2021	60-40-0	88.00	85.00	86.00	86.50
[60]	CNN	2021	70-10-20	-	-	98.70	98.47
[89]	CNN-LSTM	2022	-	**99.15**	97.60	99.62	98.38
[102]	CNN	2022	70-20-10	-	-	-	**99.30**
[94]	ResNet	2022	60-40-0	91.77	93.68	92.72	92.57
[47]	CNN	2024	5-FCV	-	-	66.35	69.42

5.2.3. Performance on PS Multi-Class Classification

PS multi-class classification is much more complex. The data cases are classified into COPD (793 cases, 86.2% of the samples), pneumonia (37 cases), healthy (35 cases), upper respiratory tract infection (23 cases), bronchiectasis (16 cases), bronchiolitis (13 cases), lower respiratory tract infection (2 cases), and asthma (1 case, 0.1% of the samples). This task requires the model to not only identify healthy individuals but also effectively differentiate between various disease types, providing detailed support for clinical diagnosis and for promoting precision medicine. Table 17 summarizes the PS multi-class recognition task of different methods.

Table 17. Performance on PS multi-class classification.

	Backbone	Year	Splitting	SEN (%)	SPE (%)	ACC (%)	HS (%)
[72]	CNN	2019	70-30-0	-	-	97.00	-
[22]	Random Forest	2020	70-30-0	-	-	88.00	87.00
[60]	CNN	2021	70-10-20	**100.00**	**98.60**	98.70	**99.30**
[136]	BiLSTM-BiGRU	2021	75-12.5-12.5	-	-	96.20	-
[107]	Bi-LSTM	2021	80-20-0	-	-	88.30	-
[31]	CNN-BiLSTM	2022	10-FCV	99.69	98.43	**99.62**	99.06
[90]	FDC-FSNet	2022	80-20-0	-	-	99.10	-
[153]	1D-CNN	2022	-	99.02	98.30	99.43	98.66
[154]	CNN-LSTM	2022	80-20-0	-	-	98.82	97.00
[89]	CNN-LSTM	2022	-	97.71	84.73	82.35	91.22
[5]	CNN	2023	50-50-0	-	-	93.00	-
[78]	VGG16	2024	60-40-0	-	-	75.00	72.00
[40]	LSTM	2024	80-20-0	99.10	89.56	94.16	94.33
[146]	CNN	2024	80-20-0	99.42	96.53	96.03	97.99

Several algorithms [31,60,153] achieve metric values larger than 98.00%. Notably, the study [60] proposes a hybrid-scale spectrogram generation method that decomposes the signals into different intrinsic mode functions and uses continuous WT for discriminative time–frequency signal representations. Additionally, a light-weight module, batch normalization, max pooling, and multi-chromatic data augmentation are embedded for accurate classification. The study [31] implements a hierarchical abstraction framework, and a Bi-LSTM-based bi-directional temporal gating mechanism is proposed to capture the pathological feature evolution in the forward and backward directions within a respiratory cycle. The study [153] combines wavelet-based denoising and Mel-frequency cepstral coefficients for feature extraction. Time-domain warping and noise injection are used to enhance data diversity; synthetic samples are adaptively generated for the minority class, and a 1D CNN is constructed for progressive temporal feature abstraction and PS multi-class prediction.

6. Discussion

After literature retrieval and screening, technical publications utilizing the ICBHI2017 database were systematically analyzed across three key aspects including signal processing, feature extraction, and sound classification. Specifically, respiratory sounds are resampled, augmented, normalized, and filtered to ensure consistency across different data sources. Quantitative features are often handcrafted in the time domain, frequency domain, and joint time–frequency domain. In addition, high-level features extracted from pre-trained deep networks have proven to be effective. Finally, the processed sounds are classified into various AS or PS categories, employing machine learning, deep learning, hybrid approaches, and other advanced learning strategies. While promising performance has been achieved on the ICBHI2017 database, there remains substantial room for RSA improvement.

6.1. The Problem of Class Imbalance

The sensitivity (SEN) value remains below 50.00% in AS classification when using the official data split (Table 13). This indicates that the models fail to correctly identify more than half of the true-positive cases, which is a serious concern in medical diagnostics. Several factors may cause this issue. First, a class imbalance plays a significant role. Normal recordings (3642 samples, 52.8%) are heavily overrepresented compared to the mixed category containing both crackle and wheeze sounds (506 samples, 7.3%). This imbalance can cause models to favor the majority class during training. Similar patterns of imbalance are also observed in related classification tasks, such as binary, ternary, and multi-class PS recognition. Second, overlapping acoustic features among crackles, wheezes, and mixed sounds complicate accurate classification, particularly when abnormalities are subtle or co-occurring. Differentiating between individual crackles or wheezes and their combination is especially difficult due to the shared intrinsic characteristics of these respiratory sounds. Third, suboptimal feature representation may limit a model's ability to learn and distinguish fine-grained patterns. While handcrafted features, features extracted from pre-trained deep networks, and hierarchical representations learned via end-to-end training have all been explored, it remains unclear which types of features are most effective for respiratory sound classification. This uncertainty makes it challenging to select the most discriminative features from the vast feature space [155]. In conclusion, the class imbalance not only induces model predictions toward the majority class but also leads to reduced sensitivity for minority classes and introduces learning bias, ultimately hindering model performance in critical clinical applications.

Advanced data augmentation methods can help address the class imbalance by generating entirely new samples. In contrast, standard augmentation techniques manipulate existing samples to promote invariant feature learning, increase data diversity, and enhance robustness to noise and distortions (Table 4). Various generative models have been proposed in the literature. Variational autoencoders (VAEs) learn to encode input data into a latent space and; then, this representation is decoded to reconstruct the original data. By optimizing a variational lower bound, VAEs enable efficient approximate inference and generative modeling [15,115,156]. Generative adversarial networks (GANs) consist of a generator to synthesize realistic samples from random noise and a discriminator to distinguish real data from generated data [157]. This adversarial setup enables models to effectively learn complex data distributions [95,158] and synthesize respiratory sounds for improved classification performance [159].

Diffusion probabilistic models generate new data by learning to reverse a gradual noising process that corrupts data over multiple steps. This involves applying a forward (diffusion) process and a reverse (generative) process iteratively [160]. One such model, DiffWave, has been used for both conditional and unconditional waveform generation [161], which have been applied to respiratory sound synthesis through adversarial fine-tuning at the Mel-spectrogram level [162,163]. Although these advanced generative models show promise, further investigation is needed to fully evaluate their effectiveness in mitigating the class imbalance in the RSA field.

By comparing performance under official data splitting (Table 13) and custom data splitting strategies (Table 14), we observe that SEN values are significantly improved when custom partitioning is applied. This demonstrates that the dataset is divided into training, validation, and testing sets and has a substantial impact on prediction performance (see Tables 15–17). In scenarios with a class imbalance, simple random splitting or standard *k*-FCV may fail to ensure fair and reliable algorithm comparisons [164]. Addressing this issue is therefore of critical importance. Several strategies can be adopted during data splitting to mitigate the class imbalance. First, stratified splitting helps preserve class distributions across all subsets, which is especially beneficial for small or multi-class datasets. Second, balancing the dataset before splitting, by down-sampling majority classes or over-sampling minority classes, can ensure more equitable representations in each subset. Third, subject-wise splitting, such as group *k*-fold, helps prevent data leakage and enhances generalization by ensuring that all samples from a single subject appear in only one partition. Additionally, if synthetic data augmentation is used, it should be performed before splitting, and care should be taken to group augmented samples with their corresponding subjects. This avoids contaminating the test set and preserves the integrity of model evaluations. However, for specific tasks with a class imbalance, the optimal application of these strategies remains unclear and warrants further investigation.

6.2. Multi-Database Fusion and Data Integration

Recent studies have increasingly adopted multi-database fusion strategies to improve the performance and generalizability of respiratory sound classification models. For instance, several works [4,51,52,63,82] have investigated the integration of the ICBHI 2017 dataset with external respiratory sound repositories such as HF_Lung and SPRSound to address challenges related to data scarcity and class imbalances.

Building upon this paradigm, the fusion of ICBHI 2017 with complementary respiratory sound datasets effectively mitigates the limitations inherent in single-source data, including an insufficient sample size, limited pathological heterogeneity, and variability in recording devices. This integration substantially enhances the robustness and generalization capability of classification models.

Specifically, the large-scale, multi-channel recordings provided by HF_Lung alongside the age-diverse and fine-grained annotations from SPRSound contribute to alleviating the class imbalance and enriching the diversity of the training corpus. Concurrently, the heterogeneity in recording equipment, sampling rates, and annotation granularity across these datasets introduces realistic domain shifts, facilitating the learning of domain-invariant representations and thereby improving model adaptability to complex clinical environments.

Consequently, multi-source data fusion paves the way for developing end-to-end diagnostic frameworks that encompass abnormal sound detection through lesion localization, establishing a robust foundation for the advancement of high-precision, clinically applicable respiratory sound analysis systems.

6.3. Feature Representation Learning

Features handcrafted in the time domain, the frequency domain, and the joint time–frequency domain are preferred (Tables 5–7), while in RSA techniques, features extracted from pre-trained deep networks are paid less attention to. To enhance the effectiveness and efficiency of deep features, more advanced foundation models could be explored, including but not limited to wav2vec [165], VGGish [166], AST [141], and masked modeling Duo [167]. These foundation models have been pre-trained with a sufficiently large and diverse audio dataset and also verified to be effective for sound analysis.

To improve the capacity of feature representation learning for respiratory sound signals, different learning strategies could be utilized. First, under the context of supervised learning, transfer learning implemented by fine-tuning a pre-trained models with a small number of labeled samples could leverage knowledge from a source task to improve the performance on a target task. Technically, a model is trained on a large dataset, such as ImageNet [168] or AudioSet [169], and then, the pre-trained model is fine-tuned on the ICBHI2017 database for AS or PS prediction [69,112,154]. Second, unsupervised learning is an emerging technique used to discover patterns or structures from unlabeled data. It learns to group the data without explicit labels [67,77]. As a type of unsupervised learning category, self-supervised learning is massively applied for training foundation models. It learns useful representations by creating pretext tasks to generate pseudo-labels, and the model is then fine-tuned for downstream tasks [77]. Contrastive learning is considered a form of self-supervised learning that designs positive pairs and negative pairs, and a model is trained to distinguish between the positive and negative pairs of the samples [67,170].

Meanwhile, various learning paradigms could be employed to improve the performance of respiratory sound classification. First, hybrid learning combines multiple learning paradigms (e.g., supervised and unsupervised or machine learning and deep learning methods) and uses complementary strengths of each paradigm to enhance robustness, accuracy, or generalization. Second, multi-modal fusion learning exploits multiple sources or types of data (modalities), such as audio, images, or clinical metadata, to improve the performance of respiratory disease detection, classification, or diagnosis. The learning paradigms can be fused at the feature level by feeding different types of features into a neural network, at the decision level by combining the outputs with weighted voting or averaging, or at the model level by fusing modality-specific feature representations through an attention mechanism or a shared latent space [148]. Third, multi-task learning integrates AS classification with PS prediction by training a single model to perform multiple tasks simultaneously. The tasks are trained together with shared layers and task-specific heads, enabling the learning of generalized features [17,66,109]. For instance, Pham et al. achieve high accuracy in both AS and PS classification tasks [66]. The well-trained model not only identifies abnormal sounds but also determines the pathological state, providing clinicians with more accurate diagnostic information.

6.4. Limitations of the Current Review

Several technical limitations remain in the current study. First, the severe class imbalance is a persistent challenge in the ICBHI2017 database. Although standard and advanced data augmentation methods have been proposed, their effectiveness remains inconclusive. Emerging approaches, such as ensemble learning, self-supervised learning, and federated learning, hold promise for addressing data scarcity and privacy constraints [171,172]. The thorough evaluation of these data augmentation strategies is essential to advance the understanding and applicability of such methods. Combining multiple databases appears to be promising for addressing the severe class imbalance, and this requires careful data processing to handle annotation inconsistencies, differences in sampling rates, variations in signal quality, and potential domain shifts. Second, no experimental benchmarking has been performed to compare the performance of the proposed classification algorithms through intra- and inter-database validation. A systematic and reproducible comparison of these techniques is highly encouraged for future work. Third, a standardized data processing pipeline is currently lacking, which hampers reproducibility and the design of fair experiments. The end-to-end RSA workflow, from signal processing to feature extraction and classification, requires consistent protocols and well-documented methodologies to support further research and industrial deployment. Fourth, while AI techniques, such as large audio and visual foundation models, have rapidly advanced in other fields, their adaptation to RSA remains underexplored. Respiratory sounds can be represented both as acoustic waveforms and Mel-spectrogram images, and large foundation models need to be tailored to accommodate these dual representations effectively. Fifth, advanced learning paradigms, including multi-modal fusion, hybrid learning, causal inference, domain adaptation, and large foundation models, offer potential solutions to the inherent limitations of relying solely on audio features for disease classification [173–176]. In addition to respiratory sounds, incorporating complementary signals, such as blood oxygen saturation, thoracic motion, electronic health record information, and other clinical exams, can enrich the model's contextual understanding [177]. Finally, the development of portable, cost-effective respiratory sound detection devices, when integrated with AI algorithms and Internet of Things technology, can significantly improve access to respiratory disease screening, continuous monitoring, and real-time processing, especially in primary care and remote settings, thereby promoting broader global adoption [171].

Some important issues should be concerned in future RSA studies. First, to strengthen methodological rigor and enhance research transparency, it is recommended to draw from established reporting guidelines such as the Preferred Reporting Items for Systematic Reviews and Meta-Analyses [178]. Adopting such frameworks can facilitate a more structured and comprehensive description of search strategies, study selection, and data extraction processes. Second, interpretability remains a critical and unresolved challenge. Understanding when and why RSA models succeed or fail is essential for improving algorithm design and clinical reliability. While interpretability is a common concern, it holds particular significance in healthcare, where transparency is vital for building clinician trust, ensuring ethical use, and supporting real-world deployment. Third, as RSA increasingly relies on open-source datasets and moves toward integrating multi-modal data, such as physiological signals, imaging, or patient metadata, issues of data privacy and security become more pressing. Ensuring compliance with data protection regulations is crucial for safeguarding patient information and enabling ethically responsible research. Fourth, regulatory and certification processes for AI-based medical devices that incorporate RSA must be considered. The successful deployment of such systems in clinical environments depends not only on algorithmic performance but also on their alignment with medical device regulations, clinical validation requirements, and usability standards. Ensuring regulatory readiness is key to translating research prototypes into safe, approved, and widely adopted clinical tools.

7. Conclusions

In this study, we present a comprehensive review of respiratory sound analysis techniques based on the ICBHI2017 database, covering signal processing, feature extraction, and classification methods. A total of 135 relevant publications are systematically analyzed. Specifically, signal processing techniques include signal resampling, data augmentation, normalization, and filtering. Feature extraction approaches span the time domain, the frequency domain, the joint time–frequency domain, and the representations from pretrained deep networks. Classification methods are categorized into machine learning, deep learning, hybrid approaches, and transformer-based models. We summarize recent advancements in classifying respiratory sounds into four AS categories under both official and custom data-splitting strategies, as well as into binary, ternary, and multi-class PS categories. The issue of the class imbalance is extensively discussed, along with strategies to mitigate its impact during data splitting and model training. Furthermore, we examine feature representation learning from various paradigms and learning strategies, and the limitations of the current review and open challenges in the field are also highlighted.

From the perspective of data sufficiency, the ICBHI2017 database remains one of the most authoritative open-source resources for respiratory sound analysis with high-quality annotation and standardized benchmarks. However, like many existing respiratory sound datasets, it suffers from a severe class imbalance, which limits algorithmic development, hinders fair performance comparison, and challenges the evaluation of model generalization. To overcome these limitations, there is a critical need to develop larger and more diverse respiratory sound databases that encompass a wider range of disease types, age groups, and health conditions. Such efforts will strengthen model generalizability, enhance adaptability to real-world clinical variability, and facilitate more reliable deployment in practical healthcare settings.

From the perspective of technical evolution, future advancements in respiratory sound analysis are expected to focus on three key directions of efficiency, generalization, and clinical applicability. A shift toward multi-modal integration and scenario-driven modeling will become increasingly prominent. To overcome the challenges related to data scarcity and high annotation costs, self-supervised learning and generative models will receive increasing attention to augment underrepresented classes of diseases. Federated learning will enable privacy-preserving and distributed model training across institutions, facilitating the integration of heterogeneous clinical data. Model architectures will be increasingly tailored to clinical requirements through the fusion of audio, physiological, and imaging modalities, enhancing diagnostic accuracy. The temporal modeling of respiratory sound sequences will support the prediction of disease progression, while interpretable frameworks will improve clinical transparency and trust. Ultimately, integrating respiratory sound analysis with omics data may pave the way toward personalized diagnostics and precision medicine.

Author Contributions: Conceptualization: S.Y., B.Z. and Q.S.; data curation: J.Y. and L.C.; formal analysis: X.L., Y.X. and Q.S.; funding acquisition: X.L., Y.X. and Q.S.; investigation: B.Z., Y.X. and Q.S.; methodology: S.Y., J.Y., B.Z. and X.L.; project administration: Q.S.; software: S.Y., J.Y. and L.C.; supervision: Q.S.; validation: X.L. and Y.X.; visualization: J.Y., L.C. and X.L.; writing—original draft: J.Y. and S.Y.; writing—review and editing: B.Z., X.L., Y.X. and Q.S. All authors have read and agreed to the published version of the manuscript.

Funding: The work was in part supported by the National Key Research and Develop Program of China (Grant Nos. 2024YFF0907401, 2022ZD0115901, and 2022YFC2409000), the National Natural Science Foundation of China (Grant Nos. 62177007, U20A20373, and 82202954), the China-Central Eastern European Countries High Education Joint Education Project (Grant No. 202012), the Application of Trusted Education Digital Identity in the Construction of Smart Campus in Vocational Colleges (Grant No. 2242000393), the Knowledge Blockchain Research Fund (Grant No. 500230), and the Medium- and Long-term Technology Plan for Radio, Television and Online Audiovisual (Grant No. ZG23011). The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Data Availability Statement: The dataset supporting the current study is available online (ICBHI 2017 Respiratory Sound Database, https://bhichallenge.med.auth.gr/ICBHI_2017_Challenge, accessed on 7 July 2025).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

RSA	Respiratory Sound Analysis
ICBHI2017	International Conference on Biomedical and Health Informatics 2017
AS	Adventitious Sound
PS	Pathological State
COPD	Chronic Obstructive Pulmonary Disease
ML	Machine Learning
DL	Deep Learning
TL	Transfer Learning
SP	Signal Processing
FE	Feature Extraction
AI	Artificial Intelligence
RMS	Root Mean Square
ZCR	Zero-Crossing Rate
FT	Fourier Transform
MFCC	Mel-Frequency Cepstral Coefficient
STFT	Short-Time Fourier Transform
WT	Wavelet Transform
CNN	Convolutional Neural Network
Bi-LSTM	Bi-directional Long Short Term Memory
AST	Audio Spectrogram Transformer
SPE	Specificity
SEN	Sensitivity
ACC	Accuracy
HS	ICBHI Score
SVM	Support Vector Machine
HMM	Hidden Markov Model
GMM	Gaussian Mixture Module
k-FCV	*k*-Fold Cross Validation
GRU	Gated Recurrent Unit
ViT	Vision Transformer
VAE	Variational Autoencoder
GAN	Generative Adversarial Network

References

1. World Health Organization. *World Health Statistics 2024: Monitoring Health for the SDGs, Sustainable Development Goals*; World Health Organization: Geneva, Switzerland, 2024.
2. Aveyard, P.; Gao, M.; Lindson, N.; Hartmann-Boyce, J.; Watkinson, P.; Young, D.; Coupland, C.A.C.; San, T.P.; Clift, A.K.; Harrison, D.; et al. Association between pre-existing respiratory disease and its treatment, and severe COVID-19: A population cohort study. *Lancet Respir. Med.* **2021**, *9*, 909–923. [CrossRef] [PubMed]
3. Xia, T.; Han, J.; Mascolo, C. Exploring machine learning for audio-based respiratory condition screening: A concise review of databases, methods, and open issues. *Exp. Biol. Med.* **2022**, *247*, 2053–2061. [CrossRef]
4. Huang, D.-M.; Huang, J.; Qiao, K.; Zhong, N.-S.; Lu, H.-Z.; Wang, W.-J. Deep Learning-Based Lung Sound Analysis for Intelligent Stethoscope. *Mil. Med Res.* **2023**, *10*, 44. [CrossRef] [PubMed]
5. Latifi, S.A.; Ghassemian, H.; Imani, M. Feature Extraction and Classification of Respiratory Sound and Lung Diseases. In Proceedings of the International Conference on Pattern Recognition and Image Analysis (IPRIA), Qom, Iran, 14–16 February 2023; pp. 1–6.
6. Sfayyih, A.H.; Sulaiman, N.; Sabry, A.H. A Review on Lung Disease Recognition by Acoustic Signal Analysis with Deep Learning Networks. *J. Big Data* **2023**, *10*, 101. [CrossRef]
7. Zarandah, Q.M.M.; Daud, S.M.; Abu-Naser, S.S. A Systematic Literature Review of Machine and Deep Learning-Based Detection and Classification Methods for Diseases Related to the Respiratory System. *Artif. Intell. Med.* **2023**, *15*, 200–215.

8. Kapetanidis, P.; Kalioras, F.; Tsakonas, C.; Tzamalis, P.; Kontogiannis, G.; Karamanidou, T.; Stavropoulos, T.G.; Nikoletseas, S. Respiratory Diseases Diagnosis Using Audio Analysis and Artificial Intelligence: A Systematic Review. *Sensors* **2024**, *24*, 1173. [CrossRef] [PubMed]
9. Altan, G.; Kutlu, Y.; Garbi, Y.; Pekmezci, A.O.; Nural, S. Multimedia Respiratory Database (RespiratoryDatabase@TR): Auscultation Sounds and Chest X-rays. *Nat. Eng. Sci.* **2017**, *2*, 59–72. [CrossRef]
10. Zhang, Q.; Zhang, J.; Yuan, J.; Huang, H.; Zhang, Y.; Zhang, B.; Lv, G.; Lin, S.; Wang, N.; Liu, X.; et al. SPRSound: Open-Source SJTU Paediatric Respiratory Sound Database. *IEEE Trans. Biomed. Circuits Syst.* **2022**, *16*, 867–881. [CrossRef]
11. Hsu, F.-S.; Huang, S.R.; Huang, C.W.; Huang, C.J.; Cheng, Y.R.; Chen, C.C.; Hsiao, J.; Chen, C.W.; Chen, L.C.; Lai, Y.C.; et al. Benchmarking of Eight Recurrent Neural Network Variants for Breath Phase and Adventitious Sound Detection on a Self-Developed Open-Access Lung Sound Database—HF_Lung_V1. *PLoS ONE* **2021**, *16*, e0254134. [CrossRef]
12. Hsu, F.-S.; Huang, S.R.; Su, C.F.; Huang, C.W.; Cheng, Y.R.; Chen, C.C.; Wu, C.Y.; Chen, C.W.; Lai, Y.C.; Cheng, T.W.; et al. A Dual-Purpose Deep Learning Model for Auscultated Lung and Tracheal Sound Analysis Based on Mixed Set Training. *Biomed. Signal Process. Control* **2023**, *86*, 105222. [CrossRef]
13. Hsu, F.-S.; Huang, S.R.; Huang, C.W.; Cheng, Y.R.; Chen, C.C.; Hsiao, J.; Chen, C.W.; Lai, F. A Progressively Expanded Database for Automated Lung Sound Analysis: An Update. *Appl. Sci.* **2022**, *12*, 7623. [CrossRef]
14. Rocha, B.M.; Filos, D.; Mendes, L.; Vogiatzis, I.; Perantoni, E.; Kaimakamis, E.; Natsiavas, P.; Oliveira, A.; Jácome, C.; Marques, A.; et al. A Respiratory Sound Database for the Development of Automated Classification. In *Precision Medicine Powered by pHealth and Connected Health*; Maglaveras, N., Chouvarda, I., De Carvalho, P., Eds.; Springer: Singapore, 2018; Volume 66, pp. 33–37.
15. Zhang, M.; Li, M.; Guo, L.; Liu, J. A Low-Cost AI-Empowered Stethoscope and a Lightweight Model for Detecting Cardiac and Respiratory Diseases From Lung and Heart Auscultation Sounds. *Sensors* **2023**, *23*, 2591. [CrossRef] [PubMed]
16. Mondal, A.; Saxena, I.; Tang, H.; Banerjee, P. A Noise Reduction Technique Based on Nonlinear Kernel Function for Heart Sound Analysis. *IEEE J. Biomed. Health Inform.* **2018**, *22*, 775–784. [CrossRef]
17. Chambres, G.; Hanna, P.; Desainte-Catherine, M. Automatic Detection of Patient with Respiratory Diseases Using Lung Sound Analysis. In Proceedings of the 2018 International Conference on Content-Based Multimedia Indexing (CBMI), La Rochelle, France, 4–6 September 2018; pp. 1–6.
18. Jakovljević, N.; Lončar-Turukalo, T. Hidden Markov Model Based Respiratory Sound Classification. In *Precision Medicine Powered by pHealth and Connected Health*; Maglaveras, N., Chouvarda, I., De Carvalho, P., Eds.; Springer: Singapore, 2018; Volume 66, pp. 39–43.
19. Demir, F.; Ismael, A.M.; Sengur, A. Classification of Lung Sounds with CNN Model Using Parallel Pooling Structure. *IEEE Access* **2020**, *8*, 105376–105383. [CrossRef]
20. Minami, K.; Lu, H.; Kamiya, T.; Mabu, S.; Kido, S. Automatic Classification of Respiratory Sounds Based on Convolutional Neural Network with Multi Images. In Proceedings of the 2020 5th International Conference on Biomedical Imaging, Signal Processing, Kitakyushu Japan, 27–29 September 2020; pp. 17–21.
21. Rocha, B.M.; Pessoa, D.; Marques, A.; Carvalho, P.; Paiva, R.P. Automatic Classification of Adventitious Respiratory Sounds: A (Un)solved Problem? *Sensors* **2020**, *21*, 57. [CrossRef]
22. Wu, L.; Li, L. Investigating into Segmentation Methods for Diagnosis of Respiratory Diseases Using Adventitious Respiratory Sounds. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 20–24 July 2020; pp. 768–771.
23. Yang, Z.; Liu, S.; Song, M.; Parada-Cabaleiro, E.; Schuller, B.W. Adventitious Respiratory Classification Using Attentive Residual Neural Networks. In Proceedings of the 21st Annual Conference of the International Speech Communication Association (INTERSPEECH 2020), Shanghai, China, 25–29 October 2020; pp. 2912–2916.
24. Asatani, N.; Kamiya, T.; Mabu, S.; Kido, S. Classification of Respiratory Sounds Using Improved Convolutional Recurrent Neural Network. *Comput. Electr. Eng.* **2021**, *94*, 107367. [CrossRef]
25. Asatani, N.; Kamiya, T.; Mabu, S.; Kido, S. Classification of Respiratory Sounds by Generated Image and Improved CRNN. In Proceedings of the 2021 21st International Conference on Control, Automation and Systems (ICCAS), Jeju, Republic of Korea, 12–15 October 2021; pp. 1804–1808.
26. Gairola, S.; Tom, F.; Kwatra, N.; Jain, M. RespireNet: A Deep Neural Network for Accurately Detecting Abnormal Lung Sounds in Limited Data Setting. In Proceedings of the 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Guadalajara, Mexico, 1–5 November 2021; pp. 527–530.
27. Gupta, S.; Agrawal, M.; Deepak, D. Gammatonegram Based Triple Classification of Lung Sounds Using Deep Convolutional Neural Network with Transfer Learning. *Biomed. Signal Process. Control* **2021**, *70*, 102947. [CrossRef]
28. Pham, L.; Phan, H.; Schindler, A.; King, R.; Mertins, A.; McLoughlin, I. Inception-Based Network and Multi-Spectrogram Ensemble Applied to Predict Respiratory Anomalies and Lung Diseases. In Proceedings of the 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Guadalajara, Mexico, 1–5 November 2021; pp. 253–256.

29. Romero Gomez, A.F.; Orjuela-Canon, A.D. Respiratory Sounds Classification Employing a Multi-Label Approach. In Proceedings of the 2021 IEEE Colombian Conference on Applications of Computational Intelligence (ColCACI), Cali, Colombia, 26–28 May 2021; pp. 1–5.
30. Stasiakiewicz, P.; Dobrowolski, A.P.; Targowski, T.; Gałązka-Świderek, N.; Sadura-Sieklucka, T.; Majka, K.; Skoczylas, A.; Lejkowski, W.; Olszewski, R. Automatic Classification of Normal and Sick Patients with Crackles Using Wavelet Packet Decomposition and Support Vector Machine. *Biomed. Signal Process. Control* **2021**, *67*, 102521. [CrossRef]
31. Fraiwan, M.; Fraiwan, L.; Alkhodari, M.; Hassanin, O. Recognition of Pulmonary Diseases From Lung Sounds Using Convolutional Neural Networks and Long Short-Term Memory. *J. Ambient Intell. Humaniz. Comput.* **2022**, *13*, 4759–4771. [CrossRef]
32. Liu, B.; Wen, Z.; Zhu, H.; Lai, J.; Wu, J.; Ping, H.; Liu, W.; Yu, G.; Zhang, J.; Liu, Z.; et al. Energy-Efficient Intelligent Pulmonary Auscultation for Post COVID-19 Era Wearable Monitoring Enabled by Two-Stage Hybrid Neural Network. In Proceedings of the 2022 IEEE International Symposium on Circuits and Systems (ISCAS), Austin, TX, USA, 27 May–1 June 2022; pp. 2220–2224.
33. Petmezas, G.; Cheimariotis, G.A.; Stefanopoulos, L.; Rocha, B.; Paiva, R.P.; Katsaggelos, A.K.; Maglaveras, N. Automated Lung Sound Classification Using a Hybrid CNN-LSTM Network and Focal Loss Function. *Sensors* **2022**, *22*, 1232. [CrossRef]
34. Ren, Z.; Nguyen, T.T.; Nejdl, W. Prototype Learning for Interpretable Respiratory Sound Analysis. In Proceedings of the ICASSP 2022—2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, 22–27 May 2022; pp. 9087–9091.
35. Tabata, M.; Lu, H.; Kamiya, T.; Mabu, S.; Kido, S. Automatic Classification of Respiratory Sound Considering Hierarchical Structure. In Proceedings of the 2022 22nd International Conference on Control, Automation and Systems (ICCAS), Busan, Republic of Korea, 27 November–1 December 2022; pp. 537–541.
36. Tong, F.; Liu, L.; Xie, X.; Hong, Q.; Li, L. Respiratory Sound Classification: From Fluid-Solid Coupling Analysis to Feature-Band Attention. *IEEE Access* **2022**, *10*, 22018–22031. [CrossRef]
37. Mang, L.D.; Canadas-Quesada, F.J.; Carabias-Orti, J.J.; Combarro, E.F.; Ranilla, J. Cochleogram-Based Adventitious Sounds Classification Using Convolutional Neural Networks. *Biomed. Signal Process. Control* **2023**, *82*, 104555. [CrossRef]
38. Papadakis, C.; Rocha, L.M.G.; Catthoor, F.; Helleputte, N.V.; Biswas, D. AusculNET: A Deep Learning Framework for Adventitious Lung Sounds Classification. In Proceedings of the 2023 30th IEEE International Conference on Electronics, Circuits and Systems (ICECS), Istanbul, Turkiye, 4–7 December 2023; pp. 1–4.
39. Crisdayanti, I.A.P.A.; Nam, S.W.; Jung, S.K.; Kim, S.-E. Attention Feature Fusion Network via Knowledge Propagation for Automated Respiratory Sound Classification. *IEEE Open J. Eng. Med. Biol.* **2024**, *5*, 383–392. [CrossRef] [PubMed]
40. Khan, R.; Khan, S.U.; Saeed, U.; Koo, I.-S. Auscultation-Based Pulmonary Disease Detection Through Parallel Transformation and Deep Learning. *Bioengineering* **2024**, *11*, 586. [CrossRef]
41. Roy, A.; Satija, U.; Karmakar, S. Pulmo-TS2ONN: A Novel Triple Scale Self Operational Neural Network for Pulmonary Disorder Detection Using Respiratory Sounds. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 1–12. [CrossRef]
42. Ren, Z.; Nguyen, T.T.; Zahed, M.M.; Nejdl, W. Self-Explaining Neural Networks for Respiratory Sound Classification with Scale-Free Interpretability. In Proceedings of the 2023 International Joint Conference on Neural Networks (IJCNN), Gold Coast, Australia, 18–23 June 2023; pp. 1–7.
43. Shi, L.; Zhang, Y.; Zhang, J. Lung Sound Recognition Method Based on Wavelet Feature Enhancement and Time-Frequency Synchronous Modeling. *IEEE J. Biomed. Health Inform.* **2023**, *27*, 308–318. [CrossRef]
44. Sun, W.; Zhang, F.; Sun, P.; Hu, Q.; Wang, J.; Zhang, M. Respiratory Sound Classification Based on Swin Transformer. In Proceedings of the 2023 8th International Conference on Signal and Image Processing (ICSIP), Wuxi, China, 8–10 July 2023; pp. 511–515.
45. Wang, F.; Yuan, X.; Meng, B. Classification of Abnormal Lung Sounds Using Deep Learning. In Proceedings of the 2023 8th International Conference on Signal and Image Processing (ICSIP), Wuxi, China, 8–10 July 2023; pp. 506–510.
46. Wu, C.; Huang, D.; Tao, X.; Qiao, K.; Lu, H.; Wang, W. Intelligent Stethoscope Using Full Self-Attention Mechanism for Abnormal Respiratory Sound Recognition. In Proceedings of the 2023 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI), Pittsburgh, PA, USA, 15–18 October 2023; pp. 1–4.
47. Cansiz, B.; Kilinc, C.U.; Serbes, G. Tunable Q-Factor Wavelet Transform Based Lung Signal Decomposition and Statistical Feature Extraction for Effective Lung Disease Classification. *Comput. Biol. Med.* **2024**, *178*, 108698. [CrossRef] [PubMed]
48. Constantinescu, C.; Brad, R.; Bărglăzan, A. Lung Sounds Anomaly Detection with Respiratory Cycle Segmentation. *Brain Broad Res. Artif. Intell. Neurosci.* **2024**, *15*, 188. [CrossRef]
49. Dexuan, Q.; Ye, Y.; Haiwen, Z.; Wenjuan, W.; Shijie, G. Classification of Respiratory Sounds Into Crackles and Noncrackles Categories via Convolutional Neural Networks. In Proceedings of the 2024 IEEE International Conference on Mechatronics and Automation (ICMA), Tianjin, China, 4–7 August 2024; pp. 800–805.
50. Hassan, U.; Singhal, A.; Chaudhary, P. Lung Disease Detection Using EasyNet. *Biomed. Signal Process. Control* **2024**, *91*, 105944. [CrossRef]

51. Song, W.; Han, J.; Deng, S.; Zheng, T.; Zheng, G.; He, Y. Joint Energy-Based Model for Semi-Supervised Respiratory Sound Classification: A Method of Insensitive to Distribution Mismatch. *IEEE J. Biomed. Health Inform.* **2024**, *29*, 1433–1443. [CrossRef]
52. Wang, F.; Yuan, X.; Bao, J.; Lam, C.-T.; Huang, G.; Chen, H. OFGST-Swin: Swin Transformer Utilizing Overlap Fusion-Based Generalized S-Transform for Respiratory Cycle Classification. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 2525913. [CrossRef]
53. Wu, C.; Ye, N.; Jiang, J. Classification and Recognition of Lung Sounds Based on Improved Bi-ResNet Model. *IEEE Access* **2024**, *12*, 73079–73094. [CrossRef]
54. Zhang, Y.; Zhang, J.; Shi, L. Open-Set Lung Sound Recognition Model Based on Conditional Gaussian Capsule Network and Variational Time-Frequency Feature Reconstruction. *Biomed. Signal Process. Control* **2024**, *87*, 105470. [CrossRef]
55. Faustino, P.; Oliveira, J.; Coimbra, M. Crackle and Wheeze Detection in Lung Sound Signals Using Convolutional Neural Networks. In Proceedings of the 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Guadalajara, Mexico, 1–5 November 2021; pp. 345–348.
56. Shokouhmand, S.; Rahman, M.M.; Faezipour, M.; Bhatt, S Abnormality Detection in Lung Sounds Using Feature Augmentation. In Proceedings of the 2023 Congress in Computer Science, Computer Engineering, & Applied Computing (CSCE), Las Vegas, Nv, USA, 24–27 July 2023; pp. 2690–2691.
57. Chen, H.; Yuan, X.; Li, J.; Pei, Z.; Zheng, X. Automatic Multi-Level In-Exhale Segmentation and Enhanced Generalized S-Transform for Wheezing Detection. *Comput. Methods Programs Biomed.* **2019**, *178*, 163–173. [CrossRef]
58. Kok, X.H.; Imtiaz, S.A.; Rodriguez-Villegas, E. A Novel Method for Automatic Identification of Respiratory Disease From Acoustic Recordings. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 2589–2592.
59. Liu, R.; Cai, S.; Zhang, K.; Hu, N. Detection of Adventitious Respiratory Sounds Based on Convolutional Neural Network. In Proceedings of the 2019 International Conference on Intelligent Informatics and Biomedical Sciences, Shanghai, China, 21–24 November 2019; pp. 298–303.
60. Shuvo, S.B.; Ali, S.N.; Swapnil, S.I.; Hasan, T.; Bhuiyan, M.I.H. A Lightweight CNN Model for Detecting Respiratory Diseases From Lung Auscultation Sounds Using EMD-CWT-Based Hybrid Scalogram. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 2595–2603. [CrossRef]
61. Zhao, Z.; Gong, Z.; Niu, M.; Ma, J.; Wang, H.; Zhang, Z.; Li, Y. Automatic Respiratory Sound Classification via Multi-Branch Temporal Convolutional Network. In Proceedings of the ICASSP 2022–2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, 22–27 May 2022; pp. 9102–9106.
62. Wang, Z.; Wang, Z. A Domain Transfer Based Data Augmentation Method for Automated Respiratory Classification. In Proceedings of the ICASSP 2022–2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, 22–27 May 2022; pp. 9017–9021. -
63. Babu, N.; Pruthviraja, D.; Mathew, J. Enhancing Lung Acoustic Signals Classification with Eigenvectors-Based and Traditional Augmentation Methods. *IEEE Access* **2024**, *12*, 87691–87700. [CrossRef]
64. Wang, Z.; Sun, Z. Performance Evaluation of Lung Sounds Classification Using Deep Learning Under Variable Parameters. *EURASIP J. Adv. Signal Process.* **2024**, *2024*, 51. [CrossRef]
65. Nguyen, T.; Pernkopf, F. Crackle Detection in Lung Sounds Using Transfer Learning and Multi-Input Convolutional Neural Networks. In Proceedings of the 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Guadalajara, Mexico, 1–5 November 2021; pp. 80–83.
66. Pham, L.; Phan, H.; Palaniappan, R.; Mertins, A.; McLoughlin, I. CNN-MoE Based Framework for Classification of Respiratory Anomalies and Lung Disease Detection. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 2938–2947. [CrossRef]
67. Song, W.; Han, J.; Song, H. Contrastive Embedding Learning Method for Respiratory Sound Classification. In Proceedings of the ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, Ontario, Canada, 6–11 June 2021; pp. 1275–1279.
68. Harvill, J.; Wani, Y.; Alam, M.; Ahuja, N.; Hasegawa-Johnsor, M.; Chestek, D.; Beiser, D.G. Estimation of Respiratory Rate From Breathing Audio. In Proceedings of the 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Glasgow, UK, 11–15 July 2022; pp. 4599–4603.
69. Lal, K.N. A Lung Sound Recognition Model to Diagnose the Respiratory Diseases by Using Transfer Learning. *Multimed. Tools Appl.* **2023**, *82*, 36615–36631. [CrossRef]
70. Wang, J.; Dong, G.; Shen, Y.; Zhang, M.; Sun, P. Lightweight Hierarchical Transformer Combining Patch-Random and Positional Encoding for Respiratory Sound Classification. In Proceedings of the 2024 9th International Conference on Signal and Image Processing (ICSIP), Nanjing, China, 12–14 July 2024; pp. 580–584.
71. Xiao, L.; Fang, L.; Yang, Y.; Tu, W. LungAdapter: Efficient Adapting Audio Spectrogram Transformer for Lung Sound Classification. In Proceedings of the 25th Annual Conference of the International Speech Communication Association (INTERSPEECH 2024), Kos Island, Greece, 1–5 September 2024; pp. 4738–4742.

72. Tariq, Z.; Shah, S.K.; Lee, Y. Lung Disease Classification Using Deep Convolutional Neural Network. In Proceedings of the 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), San Diego, CA, USA, 18–21 November 2019; pp. 732–735.
73. Nguyen, T.; Pernkopf, F. Lung Sound Classification Using Snapshot Ensemble of Convolutional Neural Networks. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 20–24 July 2020; pp. 760–763.
74. Xu, L.; Cheng, J.; Liu, J.; Kuang, H.; Wu, F.; Wang, J. ARSC-Net: Adventitious Respiratory Sound Classification Network Using Parallel Paths with Channel-Spatial Attention. In Proceedings of the 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Houston, TX, USA, 9–12 December 2021; pp. 1125–1130.
75. Yu, S.; Ding, Y.; Qian, K.; Hu, B.; Li, W.; Schuller, B.W. A Glance-and-Gaze Network for Respiratory Sound Classification. In Proceedings of the ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, 22–27 May 2022; pp. 9007–9011.
76. Zhang, L.; Lim, C.P.; Yu, Y.; Jiang, M. Sound Classification Using Evolving Ensemble Models and Particle Swarm Optimization. *Appl. Soft Comput.* **2022**, *116*, 108322. [CrossRef]
77. Kulkarni, S.; Watanabe, H.; Homma, F. Self-Supervised Audio Encoder with Contrastive Pretraining for Respiratory Anomaly Detection. In Proceedings of the 2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW), Rhodes, Greece, 4–9 June 2023; pp. 1–5.
78. Roslan, I.K.B.; Ehara, F. Detection of Respiratory Diseases From Auscultated Sounds Using VGG16 with Data Augmentation. In Proceedings of the 2024 2nd International Conference on Computer Graphics and Image Processing (CGIP), Kyoto, Japan, 12–15 January 2024; pp. 133–138.
79. Shi, L.; Zhang, J.; Yang, B.; Gao, Y. Lung Sound Recognition Method Based on Multi-Resolution Interleaved Net and Time-Frequency Feature Enhancement. *IEEE J. Biomed. Health Inform.* **2023**, *27*, 4768–4779. [CrossRef]
80. Tiwari, U.; Bhosale, S.; Chakraborty, R.; Kopparapu, S.K. Deep Lung Auscultation Using Acoustic Biomarkers for Abnormal Respiratory Sound Event Detection. In Proceedings of the ICASSP 2021—2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, Ontario, Canada, 6–11 June 2021; pp. 1305–1309.
81. Sfayyih, A.H.; Sabry, A.H.; Jameel, S.M.; Sulaiman, N.; Raafat, S.M.; Humaidi, A.J.; Kubaiaisi, Y.M.A. Acoustic-Based Deep Learning Architectures for Lung Disease Diagnosis: A Comprehensive Overview. *Diagnostics* **2023**, *13*, 1748. [CrossRef]
82. Sabry, A.H.; Dallal Bashi, O.I.; Nik Ali, N.H.; Al Kubaisi, Y.M. Lung Disease Recognition Methods Using Audio-Based Analysis with Machine Learning. *Heliyon* **2024**, *10*, e26218. [CrossRef] [PubMed]
83. Wanasinghe, T.; Bandara, S.; Madusanka, S.; Meedeniya, D.; Bandara, M.; Díez, I.D.L.T. Lung Sound Classification with Multi-Feature Integration Utilizing Lightweight CNN Model. *IEEE Access* **2024**, *12*, 21262–21276. [CrossRef]
84. Ko, T.; Peddinti, V.; Povey, D.; Khudanpur, S. Audio Augmentation for Speech Recognition. In Proceedings of the 16th Annual Conference of the International Speech Communication Association (INTERSPEECH 2015), Dresden, Germany, 6–10 September 2015; pp. 3586–3589.
85. Barbu, T. Variational Image Denoising Approach with Diffusion Porous Media Flow. *Abstr. Appl. Anal.* **2013**, 2013, 856876. [CrossRef]
86. Iqbal, T.; Helwani, K.; Krishnaswamy, A.; Wang, W. Enhancing Audio Augmentation Methods with Consistency Learning. In Proceedings of the ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, Ontario, Canada, 6–11 June 2021; pp. 646–650.
87. Aguiar, R.L.; Costa, Y.M.G.; Silla, C.N. Exploring Data Augmentation to Improve Music Genre Classification with ConvNets. In Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 8–13 July 2018; pp. 1–8.
88. Park, D.S.; Chan, W.; Zhang, Y.; Chiu, C.C.; Zoph, B.; Cubuk, E.D.; Le Q.V. SpecAugment: A Simple Data Augmentation Method for Automatic Speech Recognition. In Proceedings of the 20th Annual Conference of the International Speech Communication Association (INTERSPEECH 2019), Graz, Austria, 15–19 September 2019; p. 2613
89. Alqudah, A.M.; Qazan, S.; Obeidat, Y.M. Deep Learning Models for Detecting Respiratory Pathologies From Raw Lung Auscultation Sounds. *Soft Comput.* **2022**, *26*, 13405–13429. [CrossRef] [PubMed]
90. Tariq, Z.; Shah, S.K.; Lee, Y. Feature-Based Fusion Using CNN for Lung and Heart Sound Classification. *Sensors* **2022**, *22*, 1521. [CrossRef]
91. Wall, C.; Zhang, L.; Yu, Y.; Kumar, A.; Gao, R. A Deep Ensemble Neural Network with Attention Mechanisms for Lung Abnormality Classification Using Audio Inputs. *Sensors* **2022**, *22*, 5566. [CrossRef]
92. Chu, Y.; Wang, Q.; Zhou, E.; Fu, L.; Liu, Q.; Zheng, G. CycleGuardian: A Framework for Automatic Respiratory Sound Classification Based on Improved Deep Clustering and Contrastive Learning. *Complex Intell. Syst.* **2025**, *11*, 200. [CrossRef]

93. Rahman, M.M.; Shokouhmand, S.; Faezipour, M.; Bhatt, S. Attentional Convolutional Neural Network for Automating Pathological Lung Auscultations Using Respiratory Sounds. In Proceedings of the 2022 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 14–16 December 2022; pp. 1429–1435.
94. Nguyen, T.; Pernkopf, F. Lung Sound Classification Using Co-Tuning and Stochastic Normalization. *IEEE Trans. Biomed. Eng.* **2022**, *69*, 2872–2882. [CrossRef]
95. Kochetov, K.; Filchenkov, A. Generative Adversarial Networks for Respiratory Sound Augmentation. In Proceedings of the 2020 International Conference on Control, Robotics and Intelligent System, Xiamen, China, 27–29 October 2020; pp. 106–111.
96. Chatterjee, S.; Roychowdhury, J.; Dey, A. D-Cov19Net: A DNN Based COVID-19 Detection System Using Lung Sound. *J. Comput. Sci.* **2023**, *66*, 101926. [CrossRef] [PubMed]
97. Tariq, Z.; Shah, S.K.; Lee, Y. Multimodal Lung Disease Classification Using Deep Convolutional Neural Network. In Proceedings of the 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Seoul, Republic of Korea, 16–19 December 2020; pp. 2530–2537.
98. Im, S.; Kim, T.; Min, C.; Kang, S.; Roh, Y.; Kim, C.; Kim, M.; Kim, S.H.; Shim, K.; Koh, J.-s.; et al. Real-Time Counting of Wheezing Events From Lung Sounds Using Deep Learning Algorithms: Implications for Disease Prediction and Early Intervention. *PLoS ONE* **2023**, *18*, e0294447. [CrossRef] [PubMed]
99. Chu, Y.; Wang, Q.; Zhou, E.; Zheng, G.; Liu, Q. Hybrid Spectrogram for the Automatic Respiratory Sound Classification with Group Time Frequency Attention Network. In Proceedings of the 2023 IEEE 6th International Conference on Pattern Recognition and Artificial Intelligence (PRAI), Haikou, China, 18–20 August 2023; pp. 839–845.
100. Mahmood, A.F.; Alkababji, A.M.; Daood, A. Resilient Embedded System for Classification Respiratory Diseases in a Real Time. *Biomed. Signal Process. Control* **2024**, *90*, 105876. [CrossRef]
101. Kochetov, K.; Putin, E.; Balashov, M.; Filchenkov, A.; Shalyto, A. Noise Masking Recurrent Neural Network for Respiratory Sound Classification. In *Artificial Neural Networks and Machine Learning–ICANN 2018*; Kůrková, V., Manolopoulos, Y., Hammer, B., Iliadis, L., Maglogiannis, I., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2018; Volume 1114, pp. 208–217.
102. Borwankar, S.; Verma, J.P.; Jain, R.; Nayyar, A. Improvise Approach for Respiratory Pathologies Classification with Multilayer Convolutional Neural Networks. *Multimed. Tools Appl.* **2022**, *81*, 39185–39205. [CrossRef]
103. Jaitly, N.; Hinton, G.E. Vocal Tract Length Perturbation (VTLP) Improves Speech Recognition. In Proceedings of the ICML Workshop on Deep Learning for Audio, Speech and Language, Marseille, France, 22–23 August 2013; Volume 90, pp. 42–51.
104. Gumelar, A.B.; Yuniarno, E.M.; Anggraeni, W.; Sugiarto, I.; Mahindara, V.R.; Purnomo, M.H. Enhancing Detection of Pathological Voice Disorder Based on Deep VGG-16 CNN. In Proceedings of the 2020 3rd International Conference on Biomedical Engineering (IBIOMED), Yogyakarta, Indonesia, 6–8 October 2020; pp. 28–33.
105. Dong, G.; Wang, J.; Shen, Y.; Zhang, M.; Zhang, M.; Sun, P. Respiratory Sounds Classification by Fusing the Time-Domain and 2D Spectral Features. In Proceedings of the 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Lisbon, Portugal, 3–6 December 2024; pp. 3178–3181.
106. Wei, S.; Zou, S.; Liao, F.; Lang, W. A Comparison on Data Augmentation Methods Based on Deep Learning for Audio Classification. *J. Phys. Conf. Ser.* **2020**, *1453*, 012085. [CrossRef]
107. Yuming, Z.; Wenlong, X. Research on Classification of Respiratory Diseases Based on Multi-Features Fusion Cascade Neural Network. In Proceedings of the 2021 11th International Conference on Information Technology in Medicine and Education (ITME), Wuyishan, China, 19–21 November 2021; pp. 298–301.
108. Zhao, X.; Shao, Y.; Mai, J.; Yin, A.; Xu, S. Respiratory Sound Classification Based on BiGRU-Attention Network with XGBoost. In Proceedings of the 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Seoul, Republic of Korea, 16–19 December 2020; pp. 915–920.
109. Perna, D.; Tagarelli, A. Deep Auscultation: Predicting Respiratory Anomalies and Diseases Via Recurrent Neural Networks. In Proceedings of the 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS), Córdoba, Spain, 5–7 June 2019; pp. 50–55.
110. Manzoor, A.; Pan, Q.; Khan, H.J.; Siddeeq, S.; Bhatti, H.M.A.; Wedagu, M.A. Analysis and Detection of Lung Sounds Anomalies Based on NMA-RNN. In Proceedings of the 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Seoul, Republic of Korea, 16–19 December 2020; pp. 2498–2504.
111. Ntalampiras, S. Collaborative Framework for Automatic Classification of Respiratory Sounds. *IET Signal Process.* **2020**, *14*, 223–228. [CrossRef]
112. Ariyanti, W.; Liu, K.-C.; Chen, K.-Y.; Tsao, Y. Abnormal Respiratory Sound Identification Using Audio-Spectrogram Vision Transformer. In Proceedings of the 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Sydney, NSW, Australia, 24–27 July 2023; pp. 1–4.
113. Roy, A.; Gyanchandani, B.; Oza, A.; Singh, A. TriSpectraKAN: A Novel Approach for COPD Detection via Lung Sound Analysis. *Sci. Rep.* **2025**, *15*, 6296. [CrossRef] [PubMed]

114. Yang, R.; Lv, K.; Huang, Y.; Sun, M.; Li, J.; Yang, J. Respiratory Sound Classification by Applying Deep Neural Network with a Blocking Variable. *Appl. Sci.* **2023**, *13*, 6956. [CrossRef]
115. García-Ordás, M.T.; Benítez-Andrades, J.A.; García-Rodríguez, I.; Benavides, C.; Alaiz-Moretón, H. Detecting Respiratory Pathologies Using Convolutional Neural Networks and Variational Autoencoders for Unbalancing Data. *Sensors* **2020**, *20*, 1214. [CrossRef]
116. Rahman, M.M.; Faezipour, M.; Bhatt, S.; Vhaduri, S. AHP-CM: Attentional Homogeneous-Padded Composite Model for Respiratory Anomalies Prediction. In Proceedings of the 2023 IEEE 11th International Conference on Healthcare Informatics (ICHI), Houston, TX, USA, 26–29 June 2023; pp. 65–71.
117. Ma, Y.; Xu, X.; Li, Y. LungRN+NL: An Improved Adventitious Lung Sound Classification Using Non-Local Block ResNet Neural Network with Mixup Data Augmentation. In Proceedings of the 21st Annual Conference of the International Speech Communication Association (INTERSPEECH 2020), Shanghai, China, 25–29 October 2020; pp. 2902–2906.
118. Ma, Y.; Xu, X.; Yu, Q.; Zhang, Y.; Li, Y.; Zhao, J.; Wang, G. LungBRN: A Smart Digital Stethoscope for Detecting Respiratory Disease Using Bi-ResNet Deep Learning Algorithm. In Proceedings of the 2019 IEEE Biomedical Circuits and Systems Conference (BioCAS), Nara, Japan, 17–19 October 2019; pp. 1–4.
119. Hussin, S.F.; Birasamy, G.; Hamid, Z. Design of Butterworth Band-Pass Filter. *Politek. Kolej Komuniti J. Eng. Technol.* **2016**, *1*, 32–46.
120. Messner, E.; Fediuk, M.; Swatek, P.; Scheidl, S.; Smolle-Jüttner, F.M.; Olschewski, H.; Pernkopf, F. Multi-Channel Lung Sound Classification with Convolutional Recurrent Neural Networks. *Comput. Biol. Med.* **2020**, *122*, 103831. [CrossRef] [PubMed]
121. Levy, J.; Naitsat, A.; Zeevi, Y.Y. Classification of Audio Signals Using Spectrogram Surfaces and Extrinsic Distortion Measures. *EURASIP J. Adv. Signal Process.* **2022**, *2022*, 100. [CrossRef] [PubMed]
122. Savitzky, A.; Golay, M.J.E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* **1964**, *36*, 1627–1639. [CrossRef]
123. Rao, A.; Huynh, E.; Royston, T.J.; Kornblith, A.; Roy, S. Acoustic Methods for Pulmonary Diagnosis. *IEEE Rev. Biomed. Eng.* **2019**, *12*, 221–239. [CrossRef]
124. Kababulut, F.Y.; Kuntalp, D.G.; Düzyel, O.; Özcan, N.; Kuntalp, M. A New Shapley-Based Feature Selection Method in a Clinical Decision Support System for the Identification of Lung Diseases. *Diagnostics* **2023**, *13*, 3558. [CrossRef]
125. Seong, J.; Ortiz, B.L.; Chong, J.W. Wheeze and Crackle Discrimination Algorithm in Pneumonia Respiratory Signals. In Proceedings of the 2024 IEEE Colombian Conference on Communications and Computing (COLCOM), Barranquilla, Colombia, 21–23 August 2024; pp. 1–6.
126. Fraiwan, L.; Hassanin, O.; Fraiwan, M.; Khassawneh, B.; Ibnian, A.M.; Alkhodari, M. Automatic Identification of Respiratory Diseases From Stethoscopic Lung Sound Signals Using Ensemble Classifiers. *Biocybern. Biomed. Eng.* **2021**, *41*, 1–14. [CrossRef]
127. Kuntalp, D.G.; Özcan, N.; Düzyel, O.; Kababulut, F.Y.; Kuntalp, M. A Comparative Study of Metaheuristic Feature Selection Algorithms for Respiratory Disease Classification. *Diagnostics* **2024**, *14*, 2244. [CrossRef]
128. Yang, J.; Luo, F.L.; Nehorai, A. Spectral Contrast Enhancement: Algorithms and Comparisons. *Speech Commun.* **2003**, *39*, 33–46. [CrossRef]
129. Boggiatto, P. Landscapes of Time-Frequency Analysis: ATFA 2019. In *Applied and Numerical Harmonic Analysis Series*; Springer International Publishing AG: Cham, Switzerland, 2020.
130. Swapna, M.S.; Renjini, A.; Raj, V.; Sreejyothi, S.; Sankararaman, S. Time Series and Fractal Analyses of Wheezing: A Novel Approach. *Phys. Eng. Sci. Med.* **2020**, *43*, 1339–1347. [CrossRef]
131. Choi, Y.; Choi, H.; Lee, H.; Lee, S.; Lee, H. Lightweight Skip Connections with Efficient Feature Stacking for Respiratory Sound Classification. *IEEE Access* **2022**, *10*, 53027–53042. [CrossRef]
132. Heil, C.E.; Walnut, D.F. Continuous and Discrete Wavelet Transforms. *SIAM Rev.* **1989**, *31*, 628–666. [CrossRef]
133. Brown, J.C. Calculation of a Constant Q Spectral Transform. *J. Acoust. Soc. Am.* **1991**, *89*, 425–434. [CrossRef]
134. Perna, D. Convolutional Neural Networks Learning From Respiratory Data. In Proceedings of the 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Madrid, Spain, 3–6 December 2018; pp. 2109–2113.
135. Ntalampiras, S.; Potamitis, I. Automatic Acoustic Identification of Respiratory Diseases. *Evol. Syst.* **2021**, *12*, 69–77. [CrossRef]
136. Wall, C.; Zhang, L.; Yu, Y.; Mistry, K. Deep Recurrent Neural Networks with Attention Mechanisms for Respiratory Anomaly Classification. In Proceedings of the 2021 International Joint Conference on Neural Networks (IJCNN), Shenzhen, China, 18–22 July 2021; pp. 1–8.
137. Roy, A.; Satija, U. A Novel Multi-Head Self-Organized Operational Neural Network Architecture for Chronic Obstructive Pulmonary Disease Detection Using Lung Sounds. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2024**, *32*, 2566–2575. [CrossRef]
138. Levy, J.; Raz-Pasteur, A.; Ovics, P.; Arraf, T.; Dotan, Y.; Zeevi, Y.Y. LungNet: A Deep Learning Model for Diagnosis of Respiratory Pathologies From Lung Sounds. In Proceedings of the 2023 5th International Conference on Bio-engineering for Smart Technologies (BioSMART), Paris, France, 7–9 June 2023; pp. 1–4.

139. Bacanin, N.; Jovanovic, L.; Stoean, R.; Stoean, C.; Zivkovic, M.; Antonijevic, M.; Dobrojevic, M. Respiratory Condition Detection Using Audio Analysis and Convolutional Neural Networks Optimized by Modified Metaheuristics. *Axioms* **2024**, *13*, 335. [CrossRef]
140. Huang, G.; Liu, Z.; van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
141. Gong, Y.; Chung, Y.; Glass, J. AST: Audio spectrogram transformer. *arXiv* **2021**, arXiv:2104.01778.
142. Zou, L.; Yu, S.; Meng, T.; Zhang, Z.; Liang, X.; Xie, Y. A Technical Review of Convolutional Neural Network-Based Mammographic Breast Cancer Diagnosis. *Comput. Math. Methods Med.* **2019**, *1*, 6509357. [CrossRef]
143. Zhu, B.; Zhou, Z.; Yu, S.; Liang, X.; Xie, Y.; Sun, Q. Review of phonocardiogram signal analysis: Insights from the PhysioNet/CinC challenge 2016 database. *Electronics* **2024**, *13*, 3222. [CrossRef]
144. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
145. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
146. Shehab, S.A.; Mohammed, K.K.; Darwish, A.; Hassanien, A.E. Deep Learning and Feature Fusion-Based Lung Sound Recognition Model to Diagnose the Respiratory Diseases. *Soft Comput.* **2024**, *28*, 11667–11683. [CrossRef]
147. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
148. Neto, J.; Arrais, N.; Vinuto, T.; Lucena, J. Convolution-Vision Transformer for Automatic Lung Sound Classification. In Proceedings of the 2022 35th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), Natal, Brazil, 24–28 October 2022; pp. 97–102.
149. Moummad, I.; Farrugia, N. Pretraining Respiratory Sound Representations using Metadata and Contrastive Learning. In Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, USA, 22–25 October 2023; pp. 1–5.
150. Bae, S.; Kim, J.W.; Cho, W.Y.; Baek, H.; Son, S.; Lee, B.; Ha, C.; Tae, K.; Kim, S.; Yun, S.Y. Patch-Mix Contrastive Learning with Audio Spectrogram Transformer on Respiratory Sound Classification. In Proceedings of the 24th Annual Conference of the International Speech Communication Association (INTERSPEECH 2023) Dublin, Ireland, 20–24 August 2023; 5436–5440.
151. Kim, J.W.; Toikkanen, M.; Choi, Y.; Moon, S.E.; Jung, H.Y. BTS: Bridging Text and Sound Modalities for Metadata-Aided Respiratory Sound Classification. In Proceedings of the 25th Annual Conference of the International Speech Communication Association (INTERSPEECH 2024), Kos Island, Greece, 1–5 September 2024;1690–1694.
152. Brunese, L.; Mercaldo, F.; Reginelli, A.; Santone, A. A Neural Network-Based Method for Respiratory Sound Analysis and Lung Disease Detection. *Appl. Sci.* **2022**, *12*, 3877. [CrossRef]
153. Li, X.; Qi, B.; Wan, X.; Zhang, J.; Yang, W.; Xiao, Y.; Mao, F.; Cai, K.; Huang, L.; Zhou, J. Electret-Based Flexible Pressure Sensor for Respiratory Diseases Auxiliary Diagnosis System Using Machine Learning Technique. *Nano Energy* **2023**, *114*, 108652. [CrossRef]
154. Zhang, P.; Swaminathan, A.; Uddin, A.A. Pulmonary Disease Detection and Classification in Patient Respiratory Audio Files Using Long Short-Term Memory Neural Networks. *Front. Med.* **2023**, *10*, 1269784. [CrossRef] [PubMed]
155. Zhang, Z.; Liang, X.; Qin, W.; Yu, S.; Xie, Y. matFR: A MATLAB toolbox for feature ranking. *Bioinformatics* **2020**, *36*, 4968–4969. [CrossRef]
156. Saldanha, J.; Chakraborty, S.; Patil, S.; Kotecha, K.; Kumar, S.; Nayyar, A. Data augmentation using Variational Autoencoders for improvement of respiratory disease classification. *PLoS ONE* **2022**, *17*, e0266467. [CrossRef]
157. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), Montréal, Quebec, Canada, 8–13 December 2014; p. 27.
158. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Commun. ACM* **2020**, *63*, 139–144. [CrossRef]
159. Jayalakshmy, S.; Sudha, G.F. Conditional GAN based augmentation for predictive modeling of respiratory signals. *Comput. Biol. Med.* **2021**, *138*, 104930. [CrossRef] [PubMed]
160. Ho, J.; Jain, A.; Abbeel, P. Denoising diffusion probabilistic models. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 6840–6851.
161. Kong, Z.; Ping, W.; Huang, J.; Zhao, K.; Catanzaro, B. Diffwave: A versatile diffusion model for audio synthesis. *arXiv* **2020**, arXiv:2009.09761.
162. Han, T.T.; Le Trung, K.; Anh, P.N.; Do Trung, A. Hierarchical Embedded System Based on FPGA for Classification of Respiratory Diseases. *IEEE Access* **2020**, *13*, 93017–93032. [CrossRef]
163. Kim, J.-W.; Yoon, C.; Toikkanen, M.; Bae, S.; Jung, H.-Y. Adversarial Fine-tuning using Generated Respiratory Sound to Address Class Imbalance. *arXiv*, **2023**, arXiv:2311.06480.
164. Yu, S.; Jin, M.; Wen, T.; Zhao, L.; Zou, X.; Liang, X.; Xie, Y.; Pan, W.; Piao, C. Accurate breast cancer diagnosis using a stable feature ranking algorithm. *BMC Med. Inform. Decis. Mak.* **2023**, *23*, 64. [CrossRef] [PubMed]

165. Yuan, Y.; Xun, G.; Suo, Q.; Jia, K.; Zhang, A. Wave2vec: Deep representation learning for clinical temporal data. *Neurocomputing* **2019**, *324*, 31–42. [CrossRef]
166. Zhu, B.; Li, X.; Feng, J.; Yu, S. VGGish-BiLSTM-attention for COVID-19 identification using cough sound analysis. In Proceedings of the 2023 8th International Conference on Signal and Image Processing (ICSIP), Wuxi, China, 8–10 July 2023; pp. 49–53.
167. Niizumi, D.; Takeuchi, D.; Ohishi, Y.; Harada, N.; Kashino, K. Masked modeling duo: Towards a universal audio pre-training framework. *IEEE/ACM Trans. Audio, Speech, Lang. Process.* **2024**, *32*, 2391–2406. [CrossRef]
168. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]
169. Gemmeke, J.F.; Ellis, D.P.W.; Freedman, D.; Jansen, A.; Lawrence, W.; Moore, R.C.; Plakal, M.; Ritter, M. Audio set: An ontology and human-labeled dataset for audio events. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, New Orleans, LA, USA, 5–9 March 2017; pp. 776–780.
170. Zhang, T.; Meng, J.; Yang, Y.; Yu, S. Contrastive learning penalized cross-entropy with diversity contrastive search decoding for diagnostic report generation of reduced token repetition. *Appl. Sci.* **2024**, *14*, 2817. [CrossRef]
171. Marengo, A.; Pagano, A.; Santamato, V. An efficient cardiovascular disease prediction model through AI-driven IoT technology. *Comput. Biol. Med.* **2024**, *183*, 109330. [CrossRef]
172. Yang, T.; Yu, X.; McKeown, M.J.; Wang, Z.J. When Federated Learning Meets Medical Image Analysis: A Systematic Review with Challenges and Solutions. *Found. Trends Signal Process.* **2024**, *13*, e38. [CrossRef]
173. Suma, K.V.; Koppad, D.; Kumar, P.; Kantikar, N.A.; Ramesh, S. Multi-Task Learning for Lung Sound and Lung Disease Classification. *SN Comput. Sci.* **2024**, *6*, 51. [CrossRef]
174. Wu, Y.; Chen, J.; Hu, L.; Xu, H.; Liang, H.; Wu, J. OmniFuse: A General Modality Fusion Framework for Multi-Modality Learning on Low-Quality Medical Data. *Inf. Fusion* **2025**, *117*, 102890. [CrossRef]
175. Kursuncu, U.; Gaur, M.; Sheth, A. Knowledge Infused Learning (K-IL): Towards Deep Incorporation of Knowledge in Deep Learning. *arXiv* **2020**, arXiv:1912.00512.
176. Mokadem, N.; Jabeen, F.; Treur, J.; Taal, H.R.; Roelofsma, P.H.M.P. An Adaptive Network Model for AI-Assisted Monitoring and Management of Neonatal Respiratory Distress. *Cogn. Syst. Res.* **2024**, *86*, 101231. [CrossRef]
177. ElMoaqet, H.; Eid, M.; Glos, M.; Ryalat, M.; Penzel, T. Deep Recurrent Neural Networks for Automatic Detection of Sleep Apnea From Single Channel Respiration Signals. *Sensors* **2020**, *20*, 5037. [CrossRef] [PubMed]
178. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med.* **2009**, *6*, e1000097. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.