technologies

MDPI

Review

Advances and Challenges in Deep Learning for Acoustic Pathology Detection: A Review

Florin Bogdan * and Mihaela-Ruxandra Lascu *

Faculty of Electronics Telecommunications and Information Technologies, Politehnica University of Timisoara, 300006 Timisoara, Romania

* Correspondence: florin.bogdan@student.upt.ro (F.B.); mihaela.lascu@upt.ro (M.-R.L.)

Abstract

Recent advancements in data collection technologies, data science, and speech processing have fueled significant interest in the computational analysis of biological sounds. This enhanced analytical capability shows promise for improved understanding and detection of various pathological conditions, extending beyond traditional speech analysis to encompass other forms of acoustic data. A particularly promising and rapidly evolving area is the application of deep learning techniques for the detection and analysis of diverse pathologies, including respiratory, cardiac, and neurological disorders, through sound processing. This paper provides a comprehensive review of the current state-of-the-art in using deep learning for pathology detection via analysis of biological sounds. It highlights key successes achieved in the field, identifies existing challenges and limitations, and discusses potential future research directions. This review aims to serve as a valuable resource for researchers and clinicians working in this interdisciplinary domain.

Keywords: pathology detection; deep learning; sound processing; respiratory pathologies; cardiac pathologies; neuronal pathologies

Academic Editor: Sheryl Berlin Brahnam

Received: 17 May 2025
Revised: 24 July 2025
Accepted: 27 July 2025
Published: 1 August 2025

Citation: Bogdan, F.; Lascu, M.-R. Advances and Challenges in Deep Learning for Acoustic Pathology Detection: A Review. *Technologies* **2025**, *13*, 329. https://doi.org/10.3390/technologies13080329

1. Introduction

Traditional methods of acoustic pathology detection have long relied on the expertise of trained medical professionals to identify and diagnose various conditions based on the sounds produced by the body. These methods often involve the use of specialized equipment, such as stethoscopes and ultrasound machines, to amplify and analyze bodily sounds. The diagnosis is typically made through a subjective evaluation of these sounds, with the medical professional comparing them to a mental library of known sounds associated with specific pathologies. While these methods have proven effective in many cases, they are not without their limitations. For instance, the accuracy of diagnosis can be influenced by the experience and skill level of the medical professional, as well as the quality of the equipment used. Moreover, the diagnostic process can be time-consuming and uncomfortable for patients, particularly in cases where extensive testing is required.

In recent years, there has been a growing interest in the use of artificial intelligence (AI) to augment and enhance traditional methods of acoustic pathology detection. By leveraging the power of deep learning algorithms and advanced signal processing techniques, AI-driven pathology detection systems can analyze bio-sounds with a level of accuracy and consistency that is difficult to achieve through human evaluation alone. These systems can sift through vast amounts of data in a fraction of the time it takes a human to complete the same task, making the diagnostic process more efficient and less burdensome for patients.

So, auscultation has long offered its potential as a non-invasive method to assess an individual's physical and cognitive health status, providing insights into conditions ranging from respiratory and cardiac diseases to neurological disorders. Historically, extracting clinically relevant information from sound data has primarily involved manual feature engineering and the application of traditional machine learning methods [1].

The past decade has witnessed a transformation in acoustic signal processing driven by rapid progress in deep learning, leading to significantly improved robustness and accuracy in detecting pathological conditions from sound data [2].

This paper presents a comprehensive review of the state-of-the-art applications of deep learning for sound-based pathology detection, highlighting the significant progress achieved, alongside the challenges that researchers continue to face, as illustrated in Figure 1.

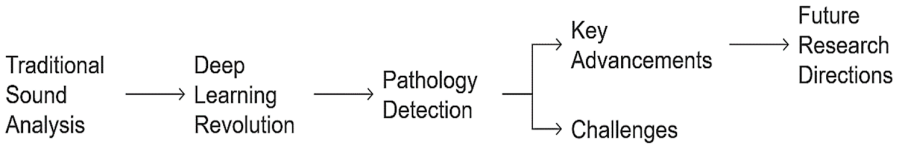

Figure 1. The "big picture" overview of this review paper—to discuss the evolution from classic pathology detection methods to deep learning approaches.

This review paper details key technical developments, the spectrum of diseases targeted, and the specific considerations pertinent to this research area. Additionally, potential future research avenues are discussed, aimed at advancing deep learning-based sound analysis towards practical diagnostic tools.

2. Methodology: Major Clinical Domains of Interest

This chapter outlines the approach and rationale behind the compilation of this review paper. It details the scope, the criteria for including and excluding literature, the justification for focusing on specific pathology areas through sound analysis, and the nature of this review as a comprehensive overview rather than a strict systematic analysis.

2.1. Review Paper Nature

It is crucial to clarify from the outset that this review adopts a narrative or standard review approach, rather than a systematic or meta-analytic review. A systematic review follows a predefined protocol to answer a specific research question by exhaustively identifying, appraising, and synthesizing all relevant studies, often culminating in quantitative analysis (meta-analysis). In contrast, this standard review provides a comprehensive and insightful overview of the state of the art by critically evaluating key developments, major trends, promising techniques, and significant challenges within the defined scope, leveraging the authors' expertise and a focused literature search strategy. This approach allows for the synthesis of diverse findings, the identification of conceptual breakthroughs, and the exploration of interconnections across different sub-fields applying the same core technology lens (deep learning) to different bio-acoustic signals.

2.2. Scope and Focus

The primary focus of this review is the application of deep learning techniques for the sound-based automatic detection and analysis of human pathologies. The review is specifically concentrated on three major categories of conditions where acoustic analysis of bio-sounds holds significant diagnostic potential:

- Cardiac Pathologies: Detection of abnormalities from heart sounds.
- Respiratory Pathologies: Detection of abnormalities from lung and breathing sounds.

- Neurological Pathologies: Detection of conditions manifesting audibly, primarily through changes in voice, speech, or other associated sounds (e.g., cough patterns).

This review encompasses studies proposing novel deep learning architectures, exploring different sound features, utilizing specific datasets, and reporting performance metrics for diagnostic tasks within these domains.

2.3. Rationale for Pathological Focus: The Significance of Cardiac, Respiratory, and Neurological Bio-Acoustics

The decision to focus on cardiac, respiratory, and neurological pathologies through the lens of sound analysis is rooted in several key factors that underscore their particular importance in the realm of bio-acoustics:

- Historical significance and accessibility: Auscultation, listening to internal body sounds, has been a foundational diagnostic technique for cardiac and respiratory conditions for centuries. These sounds are inherently accessible from the body surface using simple, non-invasive tools like stethoscopes. Neurological disorders, while not traditionally diagnosed via auscultation in the same way, often profoundly affect motor control related to speech, voice production, and even cough reflexes, creating distinct and measurable acoustic signatures.
- Direct physiological manifestation: Changes in the mechanical function of the heart, the airflow dynamics in the lungs, or the neuromuscular control of vocalization directly and instantaneously alter the characteristics of the resulting sounds. This makes acoustic analysis a direct window into the underlying physiological state.
- High disease burden: Cardiac, respiratory, and neurological disorders represent some of the leading causes of morbidity and mortality worldwide. Developing accessible, non-invasive, and potentially low-cost diagnostic tools for these areas has enormous public health implications.
- Suitability: These bio-sounds often exhibit complex, non-linear patterns, variability, and noise that traditional signal processing can struggle with. Deep Learning excels at automatically learning hierarchical features from raw or minimally processed audio data, making it particularly well-suited to capture the subtle yet complex acoustic correlates of these diverse pathologies.

While other body sounds (e.g., bowel sounds, joint sounds) are valuable for specific conditions, the combination of historical diagnostic reliance, direct physiological linkage, global health importance, and clear acoustic manifestation makes cardiac, respiratory, and neurological sounds uniquely prominent and promising targets for deep learning-based diagnosis at the current state of this research field.

2.4. Information and Literature Selection

The identification of relevant literature was an iterative process combining focused database searches with citation tracking and expert knowledge. Key academic databases (including but not limited to PubMed, IEEE Xplore, ACM Digital Library, SpringerLink, and Scopus) were systematically searched using combinations of keywords such as:

- “deep learning” OR “neural networks”;
- “cardiac sound” OR “heart sound”;
- “lung sound” OR “respiratory sound” OR “breathing sound”;
- “neurological disorder” OR “speech analysis” OR “voice analysis” OR “cough analysis”;
- “pathology detection” OR “diagnosis” OR “classification” OR “biomarker”.

Initial screening involved reviewing titles and abstracts for relevance to the core scope (deep learning, sound, pathologies listed). Full-text articles, conference papers, and reputable preprints deemed relevant were then reviewed in detail. The selection prioritized

recent publications (generally within the last 5–7 years) demonstrating novel deep learning approaches, significant dataset utilization, or high-impact results in the target areas. Citation tracking (reviewing references of key papers and identifying papers that cited them) was employed to ensure coverage of influential works and recent breakthroughs.

It is important to note that, consistent with the nature of a standard review, this process aimed for comprehensive coverage of the major advancements and prevailing themes rather than an exhaustive identification of every single publication.

2.5. Brief Inclusion of Foundational Works

While the core focus of this review is on deep learning, a limited number of foundational studies employing traditional signal processing or machine learning techniques are referenced. These works, although predating the widespread adoption of deep learning in this specific domain, are mentioned for several critical reasons:

- Context: They provide essential context for the state of the art before deep learning, outlining the challenges and limitations that earlier methods faced (e.g., reliance on manual feature engineering, limited ability to handle complex variability).
- Showing advances: By contrasting with these earlier approaches, the significant advances brought about by deep learning techniques become clearer, particularly in terms of performance, automation, and the ability to process raw or less engineered data.
- Referencing established know-how: Some foundational works introduced benchmark datasets or defined standard tasks (e.g., classifying heart sound components) that are still relevant and frequently used in deep learning research, making their inclusion necessary for understanding the evolution of the field.
- Transition to modern approaches: Including these works illustrates the intellectual trajectory from traditional analytical methods to data-driven deep learning paradigms in bio-acoustic analysis for diagnosis.

Their inclusion is selective and strictly for providing necessary historical context and demonstrating progress, not for a comprehensive review of pre-deep learning methods.

2.6. Future Significance, Contribution, and Distinction from Other Works

The methodologies employed in this review—synthesizing knowledge across specific deep learning applications in cardiac, respiratory, and neurological sound analysis—pave the way for significant future impact.

2.6.1. Future Significance

The future significance of this review lies in its role as a critical juncture, setting the path forward for the next generation of AI-powered acoustic diagnostics. By meticulously mapping the proven capabilities of deep learning across these three vital, sound-rich physiological domains, this review acts as a compass for researchers, clinicians, and technologists. It does not just summarize; it catalyzes innovation by clearly highlighting successful strategies, unresolved challenges (e.g., data scarcity, interpretability, and clinical translation), and fertile ground for future investigation. Its impact will resonate in accelerating the development of non-invasive, accessible, and potentially low-cost diagnostic and monitoring tools that could fundamentally democratize healthcare, extending expert acoustic analysis from specialized clinics to patient homes and underserved global regions alike.

2.6.2. Contribution

This review contributes more than just the synthesis of the existing literature; it offers a landscape of the state-of-the-art at a critical intersection. It forges connections and draws parallels across previously disparate domains—cardiac, respiratory, and neurolog-

ical acoustics—all viewed through the unifying and transformative revolution of deep learning. The contribution provides a panoramic yet focused view that reveals overarching trends, shared challenges, and cross opportunities. Ultimately, it provides the community with a consolidated reference that maps the exciting frontier of deep learning applied to diagnosing major diseases via sound, igniting new synergies and providing a solid foundation for further exploration and ultimately, clinical impact.

2.6.3. Distinction from Other Review Works

While other valuable reviews exist, often focusing on a single pathology type (e.g., reviews solely on heart sound analysis, or solely on lung sound analysis) or broader applications of machine learning in audio-based pathology detection, this review offers a unique triangulation. The primary distinction lies in specifically bridging three major and interconnected clinical realms—cardiac, respiratory, and neurological—through the lens of deep learning applied to their respective bio-acoustic signals. Existing reviews do illuminate corners of this space, but this work is designed to be among the first comprehensive synthesis that explicitly brings these three critical areas together under the unifying paradigm of deep learning acoustics. By comparing and contrasting approaches, challenges, and successes *across* these domains within a single framework, this review provides a holistic understanding of the field's progress and potential that is currently unavailable elsewhere, serving as a resource for researchers and clinicians interested in the full spectrum of deep learning-powered acoustic diagnostics.

2.7. Logic Organization of This Review Paper

The structure of this review paper is meticulously organized to provide a comprehensive and systematic analysis of deep learning applications in sound-based disease detection.

The paper is divided into three main sections, each dedicated to a specific pathology type: respiratory, cardiac, and neurological diseases. Within each of these sections, the discussion commences with a preamble designed to emphasize the most common diseases within that category that are amenable to sound-based detection using deep learning techniques. This is followed by a presentation of the individual research papers considered in this review, with a brief focus provided for each paper immediately after its initial mention to highlight its core contribution. Following this detailed per-paper summary, a comparison table is made to facilitate understanding of the different approaches and findings across the reviewed literature within that section. An in-depth discussion then follows, offering analysis and interpretation of the presented results from the perspective of this review article. Each section concludes with a summary highlighting several key aspects derived from the reviewed papers, including the methods employed, reported results, inherent limitations, potential future work directions, and notable advantages and disadvantages identified.

Following the comprehensive treatment of each individual pathology in this structured manner, an integrating, umbrella approach is adopted in the final sections. These sections address overarching challenges, constraints, and possible solutions pertinent across all three pathology types, culminating in final conclusions that bridge findings from all sections—a cross-sectional synthesis that represents a novel approach for a traditional review of this domain.

2.8. Applications

The application of deep learning to the field of sound processing and pathology detection has led to significant advancements in recent years. Figure 2 provides an overview. Researchers have explored a variety of deep learning architectures, including convolutional

neural networks, recurrent neural networks, and Transformers, to tackle various aspects of pathological sound analysis [3].

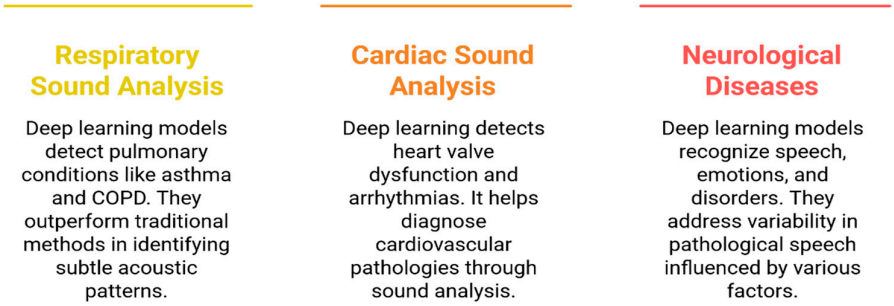

Figure 2. Medical domains of interest in sound-based pathology detection—an overview.

In the area of respiratory sound analysis, deep learning models have been trained on large datasets of lung and respiratory sounds to detect the presence of various pulmonary conditions, including asthma, chronic obstructive pulmonary disease (COPD), and pneumonia [4]. These models have demonstrated superior performance compared to traditional signal processing and machine learning approaches, with the ability to accurately identify subtle acoustic patterns indicative of respiratory abnormalities.

Similarly, with cardiac sound analysis, deep learning has been applied to the detection of heart valve dysfunction, arrhythmias, and other cardiovascular pathologies [5].

For tackling neurological diseases, deep learning models have been utilized for tasks such as automatic speech recognition, emotion recognition, and the detection of specific speech disorders, such as Parkinson's disease, dysarthria, and voice disorders. One key challenge in this field is the inherent variability in pathological speech and respiratory sounds, which can be influenced by factors such as age, gender, and severity of the condition [1].

2.9. Future Extensions and Potential Limitations

The field of sound-based pathology detection is a dynamic area, and future extensions of this review will be essential to incorporate the latest advancements. As technological barriers continue to diminish—including improved sensor quality, increased accessibility to larger datasets, and greater computational power—new and more sophisticated approaches will become feasible and necessitate inclusion. Furthermore, the rapid evolution of deep learning science, leading to the discovery of novel neural network architectures and training paradigms, will undoubtedly yield papers demonstrating improved performance, which future iterations of this review must capture. While this paper synthesizes the current state, challenges remain, particularly concerning model generalization, robustness to noise, and the need for standardized datasets. Motivated by these ongoing challenges and the promising potential of this field, the authors aim in their future work to contribute directly by designing novel neural network models capable of achieving significantly better results than the current state of the art for sound-based pathology detection.

3. Synergistic Integration of Advanced Acoustic Signal Processing and Deep Learning for Enhanced Pathology Detection

Within the application of deep learning to acoustic signal analysis, research efforts have actively investigated and implemented various techniques, including data augmentation, transfer learning, and the incorporation of domain-specific knowledge [2]. These approaches are employed to enhance model performance, improve generalization from limited data, and better capture complex patterns inherent in acoustic signals. Crucially, the field of acoustic signal processing itself has undergone significant advancements. These ad-

vancements are largely driven by the rapid evolution of computer hardware technology and the development of sophisticated sound processing software, which collectively provide the fundamental data representations necessary for effective deep learning applications. Consequently, significant progress in automated pathology detection, particularly when relying on acoustic data, is critically facilitated by the powerful synergy between sophisticated acoustic signal processing techniques and advanced deep learning methodologies [5]. This essential relationship is clearly illustrated in Figure 3.

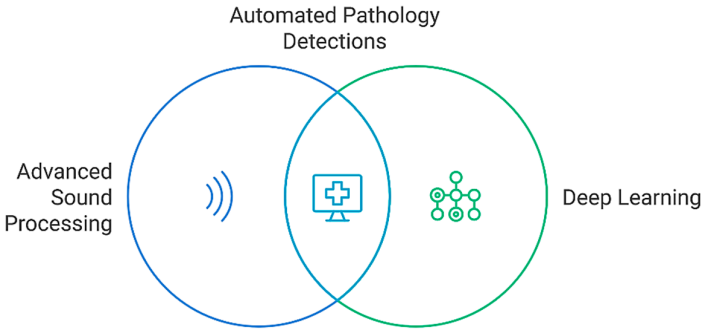

Figure 3. Futuristic approaches to pathology detection combine sound processing with deep learning.

But what sets deep learning models apart in giving good results when faced with such complicated tasks like patient diagnosis?

Firstly, deep learning models have the capacity to learn complex, non-linear representations from large, high-dimensional audio data.

Secondly, these models can effectively capture the temporal and spectral characteristics of sound signals, which are crucial for detecting subtle patterns indicative of pathological conditions.

And thirdly, the ability of deep learning to learn features automatically from raw data, without relying on manual feature engineering, has been particularly advantageous in the context of sound-based pathology detection, where the relevant acoustic signatures may not be easily identifiable by domain experts.

Basically, the usage of advanced deep learning models, such as convolutional neural networks, recurrent neural networks, and Transformers, has demonstrated their ability to extract rich, hierarchical features from audio signals, outperforming traditional signal processing and machine learning methods. Deep Learning architectures are selectively employed in audio processing based on their capacity to model specific signal properties.

For example, convolutional neural networks are well-suited for capturing time-frequency patterns in audio data, making them effective for tasks like speech recognition and music classification. Recurrent neural networks, on the other hand, excel at modeling the temporal dependencies in audio signals, enabling them to tackle problems like speech synthesis and audio event detection. More recently, transformer-based architectures have emerged as powerful tools for audio processing, showcasing their ability to capture long-range dependencies and achieve state-of-the-art performance on a variety of audio-related tasks.

4. Deep Learning for Enhanced Detection of Respiratory Pathologies

4.1. Common Aspects

Many respiratory diseases produce characteristic sounds due to airflow changes in the lungs. Here are a few examples:

- Obstructive Lung Diseases: These diseases, like asthma [6] and chronic obstructive pulmonary disease (COPD), often produce wheezing sounds due to narrowed airways.

- Restrictive Lung Diseases: Conditions like interstitial lung disease [7] can cause crackling sounds, often described as squawks, due to stiff lung tissue.
- Pneumonia: This infection can lead to crackles as well, caused by fluid buildup in the air sacs [8].

Various signal processing techniques can be applied to respiratory sounds by combining traditional and deep learning methods:

- Traditional Methods: These include spectral analysis [9] to examine sound frequency components and wavelet analysis to analyze both time and frequency information.
- Deep Learning Methods: Algorithms can be trained on large datasets of respiratory sounds to automatically classify different pathologies [10].

In Section 4 of this paper, it will be looked over six major studies, as shown in Table 1, regarding respiratory pathology detections, in a comparative and analytical manner, with the note that *studies marked below as (4), (5), and (6) referring to "squawks", are only briefly mentioned and not included in Table 1 with the deep learning models*, because, being more clinical observations, they cannot be directly compared to the deep learning studies focusing on audio analysis.

(1) *Perna (2018): "Convolutional Neural Networks Learning from Respiratory Data"* [11].

This study used convolutional neural networks (CNN) to classify respiratory sounds. The raw audio data was directly fed into the CNN model. The goal was to distinguish between normal and abnormal respiratory sounds, essentially detecting the presence of respiratory diseases.

(2) *Perna & Tagarelli (2019): "Deep Auscultation: Predicting Respiratory Anomalies and Diseases via Recurrent Neural Networks"* [10].

This study employed recurrent neural networks (RNNs) to predict respiratory anomalies and diseases from audio data. The study focused on the temporal dependencies in respiratory sounds, which RNNs are well-suited to capture. The research aimed to categorize different types of respiratory sounds indicative of specific conditions.

(3) *Shuvo et al. (2021): "A Lightweight CNN Model for Detecting Respiratory Diseases From Lung Auscultation Sounds Using EMD-CWT-Based Hybrid Scalogram"* [6].

This study proposed a lightweight CNN model for detecting respiratory diseases from lung auscultation sounds. It uses a hybrid scalogram based on Empirical Mode Decomposition (EMD) and Continuous Wavelet Transform (CWT) as input to the CNN. This preprocessing step aims to extract relevant features from the audio signals and improve the model's performance.

(4) *Pereira et al. (2019): "Squawks in Interstitial Lung Disease Prevalence and Causes in a Cohort of One Thousand Patients"* [7].

This study investigated the prevalence and causes of Interstitial Lung Disease (ILD) in a cohort of one thousand patients. It discusses the clinical presentation, diagnostic approaches, and underlying etiologies of ILD in this specific patient population.

(5) *Paciej et al. (2004): "Squawks in Pneumonia"* [8].

This study discusses a particular clinical finding or diagnostic challenge related to pneumonia.

(6) *Wodicka et al. (1997): "Respiratory Sounds"* [9].

The study refers to a survey or overview of respiratory sounds. It describes types of sounds, their physiological basis, and clinical significance.

Table 1. Comparison table of three deep learning studies in the field of respiratory pathology detection.

Feature	Perna (2018) [11]	Perna & Tagarelli (2019) [10]	Shuvo et al. (2021) [6]
Neural Network Model Type	CNN	RNN (LSTMs)	CNN (Lightweight)
Input Data	Raw Audio	Audio Data	Empirical Mode Decomposition (EMD) and Continuous Wavelet Transform (CWT)
Preprocessing	Not Specified	Not Specified	Empirical Mode Decomposition (EMD) and Continuous Wavelet Transform (CWT)
Primary Focus	Binary Classification (Normal/Abnormal)	Prediction of anomalies and diseases	Detecting Respiratory Diseases
Strengths	CNNs excel at pattern recognition	RNNs capture temporal dependencies effectively	Lightweight CNN is computationally efficient, and EMD-CWT pre-processing enhances the feature input
Limitations	May not capture temporal dynamics effectively	RNN Training can be computationally expensive	May not generalize well to unseen data

4.2. In-Depth Analysis of the Selected Papers

Both papers tackle the crucial task of automatically classifying respiratory sounds, a key step towards automated diagnosis of respiratory conditions. The common underlying principle is the application of deep learning techniques to process acoustic data derived from patient recordings. However, they diverge significantly in the specific neural network architectures employed, leading to different methodological approaches and performance characteristics.

- The CNN-based Approach [11]

This paper explores the use of convolutional neural networks (CNNs) for learning representations from respiratory data. The core idea is to transform the audio signals into a visual representation, typically a spectrogram or mel-spectrogram, where time and frequency information are represented spatially. CNNs, being highly effective in image recognition tasks, are then applied to these 2D representations. The assumption is that characteristic respiratory sounds (like crackles or wheezes) manifest as distinct patterns (textures, shapes) in the spectrogram, which CNNs are adept at identifying through convolutional and pooling layers. The work involves preprocessing respiratory sound segments into frequency-time representations. A CNN architecture is designed to learn features from these representations. This architecture typically consists of multiple convolutional layers followed by pooling layers to capture hierarchical patterns and reduce dimensionality. Finally, fully connected layers are used for classification into specific categories (e.g., the presence of crackles, wheezes, both, or normal). The focus is on leveraging the spatial feature extraction capabilities of CNNs on the spectrogram domain.

- The RNN/LSTM-based Approach [10]

This study specifically focuses on recurrent neural networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, to analyze respiratory sounds. LSTMs are designed to handle sequential data by maintaining an internal state that allows them to remember information over time. In the context of audio, this means processing the data as a sequence of acoustic features. The work involves extracting sequential features from respiratory

sound segments. An LSTM-based architecture is then employed to process these sequences. This architecture would typically involve one or more LSTM layers that process the input sequence step by step, allowing the model to learn from the temporal context of the sound events. Often, the final state of the LSTM or the output of a final pooling layer is passed to dense layers for the classification task. The objective is to exploit the ability of LSTMs to model dynamics and long-range dependencies within the respiratory sound signal.

Comparing the performance requires looking at the metrics reported in each study:

- The first study [11] demonstrated that CNNs are effective in classifying respiratory sounds based on their spectral patterns. This is because the characteristic acoustic features (like the broadband nature of crackles or the tonal quality of wheezes) create distinguishable visual patterns in the spectrogram that CNNs can learn robustly. Results show competitive accuracy and potentially strong specificity, as the CNN is good at distinguishing 'normal' patterns from 'anomalous' patterns based on learned spectral textures. However, CNNs inherently treat the spectrogram more as a static image rather than a sequence, which might slightly limit their ability to fully capture the temporal dynamics of events (e.g., the precise timing or rhythm of crackles).
- The second study [10] specifically highlights the performance advantages of RNNs, particularly LSTMs, for this task. By processing the data sequentially, LSTMs can better model the temporal evolution of the respiratory sounds. This capability is particularly beneficial for tasks where the timing and duration of sound events matter, or where context from previous time steps is important for interpreting the current one. Results from [10] show performance metrics that meet or exceed those achieved by CNNs on the same dataset. LSTMs achieve higher sensitivity (better at detecting true positives, i.e., correctly identifying anomalies/diseases), as they can better handle the nuances of how sound events unfold over time. Their ability to model temporal dependencies might also lead to improved overall classification accuracy. The "Deep Auscultation" name itself, of the study, suggests leveraging deeper temporal understanding.

In the Perna and Tagarelli paper [10], the authors used 920 annotated audio samples from 126 patients that comprise the International Conference on Biomedical Health Informatics Challenge (ICBHI) database. In the framework of a respiratory data analysis competition held in connection with the 2017 ICBHI, this competition's sound database was constructed. The audio recordings span over 5.5 h, which add up to a total of 6898 respiratory cycles, of which 1864 contain crackles, 886 contain wheezes, and 506 contain both crackles and wheezes from various respiratory pathologies.

The recordings used in the study [10] range in length from 10 to 90 s. It is known that a respiratory cycle lasts 2.7 s on average, with a standard deviation of 1.17 s; the median duration is roughly 2.54 s, while the duration varies from 0.2 s to more than 16 s.

Furthermore, wheezes have an average duration of roughly 600 ms, a minimum and maximum duration value that ranges from 26 ms to 19 s, and a relatively high variance; crackles, on the other hand, have an average duration of roughly 50 ms, a smaller variance, and minimum and maximum duration values of 3 ms and 4.88 s, respectively.

The authors have decided that an RNN (Recurrent Neural Network) should be used; the word "recurrent" implies that this kind of architecture is distinguished by applying the same action to the input sequence repeatedly. But the main characteristic that sets RNNs apart is that their output is dependent on both the input that is being used at the moment and the samples that have already been processed.

To put it another way, RNNs may find temporal connections between occurrences that are far apart in the data because they can remember knowledge about the past. The study [10] showed the performance of Long Short-Term Memory (LSTM) models and Gated

Recurrent Unit (GRU) as RNN designs, comparing them to a previous study from 2018 [11] by the same author, where CNN (Convolutional Neural Network) had been used.

Concurrently, bi-directional versions of those models were used, named "BiLSTM" [10] and "BiGRU" [10], which are distinct from their unidirectional counterparts as they integrate two hidden layers that operate in opposing directions, both contributing to the same output. This configuration allows the output layer to access information from both past (backward) and future (forward) states at the same time.

Two traditional methods of normalization were used in the study [10]: Min–Max normalization and Z-score normalization.

For the prediction tasks, the authors [10] have partitioned the ICBHI dataset into 80% for training purposes and 20% for testing. Two sets of evaluation criteria have been employed—one based on micro-averaging and one based on macro-averaging.

The authors [10] have considered:

Sensitivity for a 2-class test framework is described as follows [10]:

$$Sensitivity = \frac{C_{crackles\_or\_wheezes}}{N_{crackles\_or\_wheezes}} \quad (1)$$

Sensitivity for a 4-class test framework is described as follows [10]:

$$Sensitivity = \frac{C_{crackles} + C_{wheezes} + C_{both}}{N_{crackles} + N_{wheezes} + N_{both}} \quad (2)$$

Also, specificity has been considered as follows [10]:

$$Specificity = \frac{C_{normal}}{N_{normal}} \quad (3)$$

The values of *Cs* and *Ns* represent the count of accurately identified instances and the total count of instances, respectively, associated with the class of crackles, wheezes, both (or either crackles or wheezes) within the 4-class (or 2-class) testing framework, or normal. Similar definitions apply to the assessment of pathology-driven predictions; for example, in the context of the 3-class testing framework [10]:

$$Sensitivity = \frac{C_{chronic} + C_{non\text{-}chronic}}{N_{chronic} + N_{non\text{-}chronic}} \quad (4)$$

$$Specificity = \frac{C_{healthy}}{N_{healthy}} \quad (5)$$

Additionally, macro-averaged accuracy, precision, recall (sensitivity), and F1-score, were taken into account, where each of these metrics was calculated as the average across all classes [10]. For example, the accuracy of the 3-class pathology-driven evaluation is defined as [10]:

$$Accuracy = \frac{1}{3}\left(\frac{C_{chronic}}{N_{chronic}} + \frac{C_{non\text{-}chronic}}{N_{non\text{-}chronic}} + \frac{C_{healthy}}{N_{healthy}}\right) \quad (6)$$

Analysis of Table 2 reveals that Z-score normalization significantly enhances prediction accuracy. Relative to min–max normalization and the unnormalized baseline, Z-score scaling demonstrably achieves better predictive performance.

Table 2. Accuracy performance by LSTM models as calculated in the study by Perna and Tagarelli [10] using binary and 4-class testbeds (*maximum values in italic*).

Accuracy Performance by LSTM Models Used [10] [Binary and Four-Class Testbeds]	Un-Normalized Data		Min–Max Normalization		Z-Score Normalization	
	2-Class	4-Class	2-Class	4-Class	2-Class	4-Class
LSTM-S1	0.74	0.69	0.68	0.64	0.78	0.72
LSTM-S2	0.75	0.67	0.68	0.68	0.77	0.73
LSTM-S3	0.75	0.69	0.73	0.68	*0.81*	*0.74*
LSTM-S4	0.76	*0.70*	0.77	*0.73*	0.79	*0.74*
LSTM-S5	0.77	0.69	*0.79*	0.72	0.79	0.72
LSTM-S6	*0.78*	0.68	0.77	0.70	0.77	0.73
LSTM-S7	0.76	*0.70*	*0.79*	0.72	0.80	0.72

Upon reviewing the outcomes of the binary test framework presented in Table 3, it is evident that the LSTM-based approaches yield the most favorable performance. Notably, the frame composition configurations S4 and S7 enables the surpass of the CNN-based method, achieving improvements of up to 16% in accuracy, 9% in recall, 6% in F1-score, 4% in specificity, and 3% in sensitivity.

Table 3. Performance of the LSTM-based methods, as calculated in the study by Perna and Tagarelli [10] vs. CNN-based method as calculated in the previous study by Perna [11] using 2-class testbeds (*maximum values in italic*).

No. of Classes Used	LSTM Models Used in the Study [10] vs. CNN Model Used in the Previous Study [11]	Accuracy	Precision	Recall	F1-Score	Specificity	Sensitivity
2	CNN	0.83	*0.95*	0.83	0.88	0.78	0.97
2	LSTM-S1	0.98	0.92	0.85	0.88	0.70	*1.00*
2	LSTM-S3	0.98	0.93	0.87	0.89	0.77	0.99
2	LSTM-S4	*0.99*	*0.95*	*0.92*	*0.94*	0.79	*1.00*
2	LSTM-S6	0.98	0.92	0.88	0.90	0.80	0.99
2	LSTM-S7	*0.99*	0.94	0.91	0.92	*0.82*	0.99
3	CNN	0.82	0.87	0.82	0.84	0.76	0.89
3	LSTM-S1	0.97	0.91	0.88	0.89	0.75	0.97
3	LSTM-S3	0.97	0.92	0.88	0.90	0.80	*0.98*
3	LSTM-S4	*0.98*	0.91	*0.90*	0.90	0.80	*0.98*
3	LSTM-S6	0.97	0.91	0.87	0.89	*0.82*	*0.98*
3	LSTM-S7	*0.98*	*0.93*	*0.90*	*0.91*	*0.82*	*0.98*

In another study, the neural network proposed by Shuvo et al., from 2021 [6], shows a clear accuracy improvement vs. a classic VGG16 neural network model. The study presents a novel approach for classifying respiratory diseases from lung auscultation sounds, addressing the challenge of noise and variability inherent in such audio data and the need for efficient deployment. The core methodology revolves around a two-stage process: generating a robust time-frequency representation of the audio using a hybrid Empirical Mode Decomposition (EMD) and Continuous Wavelet Transform (CWT) approach, followed by

classification using a specially designed lightweight Convolutional Neural Network (CNN). The description of work involves processing raw lung sound recordings, sourced from the public ICBHI 2017 Challenge dataset, through this hybrid feature extraction step to create visual "scalogram" inputs for the neural network. The lightweight CNN architecture is then trained on these scalogram images to distinguish between different respiratory states, including normal, wheezing, crackling, and potentially combinations thereof, aiming for high accuracy while minimizing computational complexity.

The key methodological innovation lies in the pre-processing stage. Instead of relying solely on standard spectrograms or scalograms, the authors propose a hybrid scalogram generation technique. This involves first decomposing the audio signal into its Intrinsic Mode Functions (IMFs) using EMD, which helps in separating different signal components based on their oscillation characteristics.

Subsequently, the CWT is applied to these individual IMFs, generating time-frequency representations for each component. These representations are then combined or stacked to form a multi-channel "hybrid scalogram" image. This intricate feature extraction aims to produce a more informative and potentially noise-resistant visual representation of the lung sounds than traditional methods, highlighting subtle pathological patterns that might be missed otherwise. This processed visual feature serves as the input to the classification model.

The classification is performed by a lightweight CNN model. The "lightweight" aspect is crucial, suggesting an architecture designed with fewer parameters and computational operations compared to larger, state-of-the-art CNNs. This design choice explicitly targets feasibility for deployment on resource-constrained devices, such as mobile phones or portable diagnostic equipment, making the technology more accessible. The CNN learns hierarchical features from the hybrid scalogram images, ultimately outputting a classification decision regarding the presence and type of respiratory abnormality. The results involve performance metrics like accuracy, sensitivity, specificity, and composite score. The paper demonstrates that the proposed method, combining the effective EMD-CWT hybrid scalogram features with the efficient lightweight CNN, achieves competitive or state-of-the-art performance while maintaining a significantly reduced model size and computational footprint compared to existing, more complex deep learning methods, thereby validating both the feature extraction technique and the practical utility of the lightweight model, and proving that the Lightweight CNN Model is more efficient than classic VGG 16 model, as shown in Table 4.

Table 4. Results from the 2021 study by Shuvo and others [6] performed on the same ICBHI 2017 dataset vs. a classic VGG16 model (*maximum values in italic*).

Models Used in the Shuvo et al. Study [6]	Chronic Classification				Pathological Classification			
	Precision	**Recall**	**Accuracy**	**F1 Score**	**Precision**	**Recall**	**Accuracy**	**F1 Score**
Proposed Lightweight CNN Model [6]	*98.9*	*98.9*	*98.92*	*98.9*	*98.68*	*98.27*	*98.7*	*98.47*
Classic VGG16 Model [6]	97.95	97.83	97.84	97.89	97.6	97.86	97.62	97.01

4.3. Observations from Section 4

To fully appreciate the practical utility of deep learning in the detection and classification of sound-based respiratory pathologies, it is crucial to first delineate their clinical benefits. These advantages are schematically presented in Figure 4.

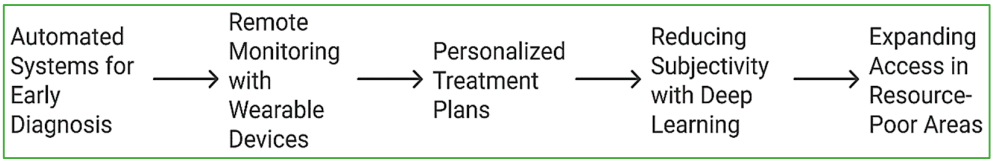

Figure 4. Advantages of using deep learning in sound-based respiratory pathologies detection and classification.

Automated systems notably contribute to the earlier diagnosis of respiratory diseases, such as asthma and chronic obstructive pulmonary disease (COPD), thereby facilitating prompt therapeutic intervention. Furthermore, wearable devices enable remote monitoring of respiratory health, proactively alerting both patients and clinicians to emergent issues. The application of sound analysis further permits the development of personalized treatment plans, tailored precisely to individual patient responses.

Deep learning models inherently offer a more objective and consistent analysis of respiratory sounds, significantly mitigating the subjectivity pervasive in human expert evaluations. Beyond these benefits, automated systems hold substantial promise for expanding access to sophisticated respiratory sound analysis, particularly in resource-limited regions where specialized medical personnel are scarce.

5. Deep Learning for Automated Analysis of Cardiac Pathology

5.1. Common Aspects

While the majority of the research in this field has focused on respiratory pathologies, there is also growing interest in applying deep learning techniques to the analysis of cardiac sounds for the detection of various heart-related conditions.

Cardiac auscultation, the process of listening to heart sounds, has long been an important diagnostic tool [5,12]. The ability to accurately detect and classify heart sounds, such as murmurs, can provide valuable insights into the presence and severity of heart diseases.

For example, studies have demonstrated the effectiveness of convolutional neural networks in identifying heart valve diseases, arrhythmias, and other cardiac abnormalities from heart sound recordings [12], and have reviewed the application of deep learning to heart sound analysis, highlighting the significant progress made in the past few years and the potential for further advancements in this field.

Examples of heart sounds that can be identified as heart diseases using deep learning are as follows:

- Murmurs: Abnormal heart sounds caused by turbulent blood flow, which can indicate valve problems or other structural issues.
- Arrhythmias: Irregular heartbeat patterns, which could signal conditions like atrial fibrillation or ventricular tachycardia.
- Stenosis: Narrowing of heart valves, which can cause distinctive sounds.
- Cardiomyopathy: Thickening or weakening of the heart muscle, which may produce altered heart sounds.

As with respiratory diseases, deep learning goes hand in hand with advanced sound processing techniques, can significantly enhance the early detection, monitoring, and personalization of treatment for a wide range of cardiac pathologies [5,13].

Section 5 of this paper looks over four major studies regarding cardiac pathology detections, in a comparative and analytical manner, as shown in Table 5.

Table 5. Comparison table of the four chosen studies for cardiac pathologies.

Feature	Reed et al. (2004) [12]	Chorba et al. (2021) [14]	Sfayyih et al. (2023) [13]	Ren et al. (2024) [5]
Study focus	General review of heart sound analysis techniques	Deep learning for cardiac murmur detection	Deep learning for acoustic-based lung disease diagnosis	Survey of deep learning in heart sound analysis
Methodology used	Review of existing signal processing and classification	CNN-based deep learning algorithm	Review of deep learning architectures	Comprehensive survey of DL methods
Data source	Not specified (general review)	Digital stethoscope recordings of heart sounds	Not specified (general review)	Not applicable (survey)
Target pathologies	Heart disease diagnosis	Cardiac murmur detection	Lung disease diagnosis	All aspects of heart sound analysis using DL
Neural network model type	Traditional signal processing, feature extraction, and classification	Convolutional Neural Network (CNN)	CNN, RNN, Transformers	DL models, including CNN, RNN, Transformers, etc.
Key findings and results	Review of foundational methods and challenges; no specific accuracy.	Demonstrated high accuracy in cardiac murmur detection	Highlights the success of DL in lung disease diagnosis	Improvement in accuracy and efficiency for detection using DL highlights the need for better model construction
Limitations/Challenges Highlighted	Variability in heart sounds, noise, and need for preprocessing	Not explicitly stated in the abstract; Likely dataset size and generalization.	Data variability, dataset size, noise sensitivity.	Explainability, robustness to noise, and data scarcity.

(1) *Reed, T., Reed, E. N., and Fritzson, P. (2004) "Heart sound analysis for symptom detection and computer-aided diagnosis"* [12].

The paper presents a comprehensive review of heart sound analysis techniques for detecting symptoms and aiding in computer-assisted diagnosis. It explores various signal processing methods used to extract features from heart sounds, including time-domain, frequency-domain, and time-frequency domain techniques. These features are then used to classify different heart conditions. The study emphasizes the importance of preprocessing steps like noise reduction and segmentation. It also discusses the challenges and limitations of heart sound analysis, such as the variability in heart sounds and the presence of noise. The work, generally, provides a foundational review of the field at the time.

(2) *Chorba JS. et al. (2021) "Deep Learning Algorithm for Automated Cardiac Murmur Detection via a Digital Stethoscope Platform"* [14].

The study focuses on developing a deep learning algorithm for automated cardiac murmur detection utilizing digital stethoscope recordings. The researchers trained a CNN on a large dataset of heart sound recordings, including both normal heart sounds and those with various murmurs. The algorithm was designed to classify heart sounds as either containing a murmur or being normal. The study demonstrates the potential of deep learning to improve the accuracy and efficiency of cardiac murmur detection, potentially leading to earlier diagnosis and treatment of heart conditions. The use of a digital stethoscope platform makes the technology more accessible and practical for widespread use.

(3) *Sfayyih, H. A. et al. (2023) "Acoustic-Based Deep Learning Architectures for Lung Disease Diagnosis: A Comprehensive Overview"* [13].

The paper provides a comprehensive overview of the use of deep learning architectures for lung disease diagnosis based on acoustic analysis (primarily cough and breath sounds). It reviews various deep learning models, including CNNs, RNNs, and Transformers, and their application to classifying different lung conditions such as pneumonia, asthma, and COPD. The study highlights the advantages of deep learning in automatically extracting relevant features from acoustic signals and achieving high accuracy in disease classification. It also discusses the challenges associated with acoustic-based lung disease diagnosis, such as noise variability in recording conditions, and the need for large, labeled datasets.

(4) *Ren, Z. et al. (2024) "A Comprehensive Survey on Heart Sound Analysis in the Deep Learning Era"* [5].

The paper presents a comprehensive survey of heart sound analysis techniques in the context of deep learning advancements. It reviews various deep learning approaches used for different tasks, including heart sound classification, segmentation, and anomaly detection. The paper discusses the strengths and weaknesses of different deep learning models in this domain, highlighting the importance of factors like dataset size, data augmentation, and model architecture. It also identifies current challenges and future research directions, such as the need for explainable deep learning and the development of robust models that can handle noisy data. The study also emphasizes the significant impact of deep learning on improving the accuracy and efficiency of heart sound analysis.

5.2. In-Depth Analysis of the Chosen Papers

For this discussion, a narrowing comparison between two of the most notable studies in this field, by Reed et al. [12] from 2004 and by Ren et al. [5] from 2024, will be made.

Being 20 years apart, it will be relevant to show the significant approach changes that have occurred in this wide timeline, and it will also be a great opportunity to highlight the improvements that have arisen due to major technological breakthroughs.

It is important to remember that these papers have drastically different goals and scopes. The first one [12] from 2004 is a research paper focused on methods, while the more recent one [5], from 2024, is a comprehensive survey covering nearly two decades of advancements. Therefore, the comparison will focus on key aspects of their approaches and perspectives.

Table 6 summarizes the key differences based on the provided information and common knowledge from the two above-mentioned publications.

Table 6. Heart sound analysis approach comparison table over a wide, twenty-year time period.

Feature	Reed et al. (2004) [12]	Ren et al. (2024) [5]
Study type	Research Article (focused on specific methods)	Survey Paper (Comprehensive overview)
Study scope	Specific signal processing and machine learning techniques for heart sound analysis.	Broad coverage of deep learning methods in heart sound analysis, including preprocessing, segmentation, classification, and emerging trends.
Time period	Focus on techniques and research prevalent *before* 2004.	Focus on developments *since* around 2000s, with emphasis on the deep learning era (2010s onwards).

Table 6. *Cont.*

Feature	Reed et al. (2004) [12]	Ren et al. (2024) [5]
Methodology focus	Traditional signal processing (e.g., time-frequency analysis, wavelet transforms), and early machine learning (e.g., rule-based systems, basic neural networks).	Primarily deep learning architectures (CNNs, RNNs, Transformers), with some discussion of pre-deep learning methods for context.
Emphasis	Feature extraction, algorithm development for specific heart sound characteristics, and rule-based systems.	Automated feature learning, end-to-end models, large-scale datasets, and performance benchmarks.
Dataset size	Smaller, using standard benchmark datasets available at the time.	Discussion of very large datasets and the challenges/opportunities they present for deep learning.
Hardware/Computational resources	Limited computational resources compared to 2024. Algorithms needed to be efficient.	Assumes availability of significant computational resources (like advanced graphics hardware) to train large deep learning models.
Example of techniques discussed	Time-frequency analysis, Wavelet transforms, Rule-Based Systems, Basic Neural Networks	CNNs, RNNs, LSTMs, GRUs, Transformers, Attention Mechanisms, Transfer Learning, Federated Learning
Overall goal	To develop and evaluate specific algorithms for heart sound analysis for symptom detection and computer-aided diagnosis.	To provide a comprehensive overview of the landscape of deep learning techniques applied to heart sound analysis, highlighting advancements, challenges, and future research directions.
Impact discussed with deep learning techniques	Not Applicable	Transfer learning from other fields like speech processing
Coronary Artery Disease focus	Focuses of rule-based systems and knowledge-based approach for coronary artery disease	Focuses on DL algorithm performance for coronary artery disease

Table 6 highlights several key differences between heart sound analysis research in 2004 and 2024. The most notable change is a technological shift, with deep learning techniques now dominating the field, contrasting with the reliance on traditional signal processing and machine learning methods in 2004. Furthermore, the scope of research has broadened significantly; the earlier work focused on specific research investigations, whereas more recent studies offer encompassing overviews of the entire field. Deep learning has also automated feature extraction, a task that previously required manual engineering. This advancement, along with the availability of large datasets and the rise in computational power, has greatly fueled the deep learning revolution in heart sound analysis.

Very good results have also been obtained from a study by Chorba JS. et al., from 2021 [14] as shown in Table 7. The study used auscultation recordings. The authors [14] have developed and tested a deep learning algorithm for automated cardiac murmur detection using a digital stethoscope platform. Their study demonstrated that the algorithm, trained on a large dataset of heart sounds, achieved high accuracy in identifying the presence or absence of murmurs. This suggests the potential for using this technology to screen for heart conditions in a convenient and accessible way, particularly in settings where expert auscultation is limited. The study supports the feasibility of using AI-powered digital stethoscopes for automated cardiac murmur detection.

However, the paper by Ren et al. from 2024 [5], provides a broader, comprehensive overview of heart sound analysis techniques using deep learning. This study [5] delves

into multiple facets of heart sound analysis, beginning with an examination of diverse deep learning architectures, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and Transformers. It highlights the advantages and disadvantages of each architecture in performing tasks like heart sound classification, segmentation, and anomaly detection. The research further discusses the existing datasets suitable for heart sound analysis and the prevalent preprocessing methods employed to improve the quality of heart sound signals before they are used in deep learning models. The paper also reviews the practical applications of deep learning-based heart sound analysis in the diagnosis of different cardiac conditions, including valvular heart disease, coronary artery disease, and congenital heart defects. Finally, the authors [5], pinpoint significant challenges within the field, such as the scarcity of high-quality, annotated data, the necessity for resilient models capable of generalizing across diverse patient demographics and recording scenarios, and the critical role of Explainable AI (XAI) in understanding the predictions made by deep learning models. Building upon these challenges, the study [5] proposes potential avenues for future research aimed at overcoming these obstacles. In essence, the paper [5] demonstrates that deep learning (DL) is a powerful tool for heart sound analysis, but also emphasizes the remaining hurdles and opportunities for improvement in the field. Deep learning (DL) has shown superior proficiency in the analysis of heart sounds compared to traditional machine learning (ML) methods. DL models generally utilize raw audio signals or time-frequency representations as their input, which enhances efficiency by eliminating the necessity for manually selecting acoustic features. Additionally, the intricate architectures of DL models shows their capacity to extract abstract representations from extensive datasets.

Table 7. Results from the study Chorba JS. et al. (2021) [14].

Heart Sounds	Recordings of Murmur (Total)	Sensitivity	Specificity
	Annotated grade		
All murmurs	499	76.3	91.4
	Position		
Aortic	146	75	89.4
Mitral	126	71.2	92.7
Pulmonic	189	81.9	91.1
Tricuspid	113	75.4	92.4

In Table 8, the performance results of using different neural network architectures for heart sound analysis are shown, as identified in the study [5]. To implement deep learning algorithms for sound analysis effectively, it is essential to sequentially apply the processes of denoising, segmentation, classification, and interpretation, each utilizing various methodological approaches. Systematic evaluation of these approaches is critical to identify the optimal combination that maximizes performance outcomes.

Table 8. Performance results of using different neural network architectures for heart sound analysis in the study by Ren et al., from 2024 [5].

Architecture	Typical Performance Range (Accuracy or F1-Score)	Notes
Convolutional Neural Networks (CNNs)	75–95%	Effective for feature extraction directly from raw or pre-processed heart sound signals (e.g., spectrograms). Performance depends heavily on network depth, filter size, and training data size. 1D-CNNs are commonly used.

Table 8. *Cont.*

Architecture	Typical Performance Range (Accuracy or F1-Score)	Notes
Recurrent Neural Networks (RNNs)/LSTMs/GRUs	70–92%	Well-suited for capturing temporal dependencies in heart sound signals. LSTMs and GRUs address the vanishing gradient problem in standard RNNs, enabling them to learn long-range dependencies. Often used in combination with CNNs.
Convolutional Recurrent Neural Networks (CRNNs)	80–97%	Combines the feature extraction capabilities of CNNs with the temporal modeling abilities of RNNs. Generally achieves higher performance than CNNs or RNNs alone.
Attention Mechanisms (e.g., Transformers)	85–98%	Attention mechanisms allow the model to focus on the most relevant parts of the heart sound signal. Transformers, in particular, have shown state-of-the-art performance in various sequence modeling tasks and are increasingly being applied to heart sound analysis.
Ensemble Methods	82–99%	Combining multiple models (e.g., CNNs, RNNs, and/or traditional machine learning classifiers) can improve robustness and accuracy. Diversity in the ensemble is key.
Other Advanced Architectures (e.g., Graph Neural Networks, Autoencoders)	Varies Widely (Limited Data in Survey)	These architectures are less commonly reported in the survey but show promise for specific tasks, such as anomaly detection or representation learning. Performance is highly dependent on the specific architecture and application. GNNs are useful for multi lead heart sounds.

5.3. Observations from Section 5

Deep learning (DL) offers significant clinical advantages for identifying cardiac pathologies, primarily by demonstrably improving the accuracy of heart sound detection and classification, as seen in the Pros–Cons diagram from Figure 5. Furthermore, DL-powered automated systems enable real-time patient monitoring and diagnosis, thereby alleviating the workload of healthcare professionals. This automation also imparts a notable degree of objectivity, ensuring a more consistent and unbiased analysis compared to human interpretation. Consequently, DL holds potential for broadening access to advanced heart sound analysis, particularly in underserved regions with limited specialized medical expertise. These advancements collectively underscore a significant potential for the early detection of cardiac abnormalities, which, in turn, can facilitate timely interventions and improve patient outcomes.

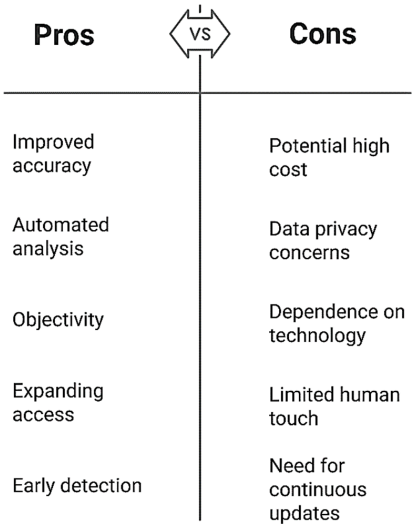

Figure 5. Pros and cons—a short summary regarding deep learning respiratory pathologies detection.

6. Deep Learning in the Diagnosis of Neurological Disorders

6.1. Common Aspects

Many neurological conditions, such as Parkinson's disease, Alzheimer's disease, and Amyotrophic lateral sclerosis, can manifest in subtle changes in a person's speech patterns and voice characteristics [15,16].

The application of deep learning techniques has enabled the identification of subtle acoustic characteristics associated with neurological disorders, potentially facilitating early detection and improved patient management.

According to a review article from 2018 [2], the field of speech analysis for health applications has seen a growing impact of deep learning techniques. The review provided an overview of the current state-of-the-art in this domain, highlighting how deep learning-based solutions are becoming more prevalent in the literature. The article noted that while deep learning has not had the same overall dominating effect as in other fields, in the medical field of neurological disorders there are substantial and ongoing research efforts exploring the use of intelligent signal analysis and deep learning techniques to extract various facets of information from speech signals, such as linguistic content, paralinguistic states, and speaker traits, with the aim of robust and accurate patterns recognition.

Clinical benefits of using deep learning for neurological pathology identification through sound analysis include

- Parkinson's Disease (PD): Studies have demonstrated the ability of deep learning models to detect subtle changes in speech patterns, such as reduced vocal intensity, increased hoarseness, and irregular articulation, which can be indicative of Parkinson's disease [17].
- Alzheimer's Disease (AD): Deep learning has been applied to analyze speech and language features, such as word usage, syntax, and semantic content, to identify early signs of cognitive decline associated with Alzheimer's disease [18,19].
- Amyotrophic Lateral Sclerosis (ALS): Changes in speech, voice, and swallowing are common in ALS, and deep learning techniques have been employed to analyze these acoustic features for early detection and monitoring of the disease [20–22].

Key advantages of using deep learning for neurological pathology identification through sound analysis include aspects like the fact that deep learning facilitates early detection by recognizing subtle acoustic changes that may appear before obvious clinical symptoms, leading to earlier diagnosis and intervention. Additionally, these automated systems provide a more objective assessment compared to subjective human evaluations, ensuring greater consistency and reduced bias. The scalability of deep learning algorithms also makes them well-suited for widespread deployment, potentially expanding access to neurological screening and monitoring, particularly in areas with limited resources. Furthermore, these systems can be seamlessly integrated into wearable devices or mobile apps, allowing for continuous monitoring and timely detection of disease progression or relapse. Also, deep learning models can be customized to individual patient characteristics, paving the way for more personalized treatment and management approaches.

Section 6 of this paper looks over nine major studies regarding neurological pathology detections, in a comparative and analytical manner, as shown in Table 9.

(1) *Cummins, N., Baird, A. and Schüller, B. (2018) "Speech analysis for health: Current state-of-the-art and the increasing impact of deep learning"* [2].

The article provides a broad overview of the field of speech analysis for health applications. It examines the current state-of-the-art, covering various speech features (e.g., acoustic, prosodic, and articulatory), machine learning techniques used for classification

and prediction, and the clinical applications addressed (e.g., depression, Alzheimer's disease, and Parkinson's disease). A key focus is on the increasing impact of deep learning methods on speech analysis for health. The authors discuss the advantages, limitations, and challenges associated with deep learning models, such as the need for large datasets and the interpretability of the models. They also touch on future directions in the field, including the development of more robust and personalized speech-based diagnostic tools. The paper emphasizes the potential of speech as a non-invasive and cost-effective biomarker for a wide range of health conditions.

(2) *Vizza, P. et al. (2019) "Methodologies of speech analysis for neurodegenerative diseases evaluation"* [16].

The paper focuses specifically on methodologies used for speech analysis in the context of neurodegenerative diseases. It reviews the different stages involved in speech-based evaluation, including speech data acquisition (elicited vs. spontaneous speech), pre-processing techniques (noise reduction and voice activity detection), feature extraction (acoustic, prosodic, and articulatory), and classification algorithms (e.g., Gaussian Mixture Models). The authors provide a detailed overview of the speech characteristics that are typically affected by various neurodegenerative diseases like Parkinson's disease, Alzheimer's disease (AD), and Amyotrophic Lateral Sclerosis (ALS). The paper highlights the challenges and opportunities in using speech to detect early signs of these diseases and track their progression. The importance of standardized data collection and analysis protocols is also emphasized.

(3) *Ipiña, d, L. K. et al. (2020) "On the analysis of speech and disfluencies for automatic detection of Mild Cognitive Impairment"* [18].

The study investigates the use of speech and disfluency analysis for the automatic detection of Mild Cognitive Impairment (MCI). The authors hypothesize that individuals with MCI exhibit specific patterns of speech disfluencies (e.g., hesitations, repetitions, fillers) and alterations in speech characteristics compared to healthy controls. The study involves collecting speech samples from MCI patients and healthy individuals, extracting acoustic and disfluency-related features, and training machine learning models to classify individuals as either MCI or healthy. The paper presents experimental results demonstrating the effectiveness of the proposed approach in detecting MCI based on speech analysis. The findings suggest that speech disfluencies can serve as a valuable biomarker for early detection of cognitive decline.

(4) *Teplansky et al. (2023) "Measuring Articulatory Patterns in Amyotrophic Lateral Sclerosis Using a Data-Driven Articulatory Consonant Distinctiveness Space Approach"* [21].

The paper introduces a data-driven method for analyzing articulatory patterns in individuals with Amyotrophic Lateral Sclerosis (ALS). An "articulatory consonant distinctiveness space" is developed to measure and visualize changes in speech production, providing a quantitative approach to track articulatory decline in ALS patients. This new clinic approach aims to offer a more sensitive and detailed understanding of speech impairments associated with the disease compared to traditional methods.

(5) *Stegmann, G. et al. (2020) "Early detection and tracking of bulbar changes in ALS via frequent and remote speech analysis"* [22].

Building upon previous work, the study focuses on the early detection and tracking of bulbar changes in ALS patients through frequent and remote speech analysis. The authors propose a system that enables patients to record speech samples regularly from their homes, allowing for continuous monitoring of their speech characteristics. The system analyzes these speech samples to detect subtle changes in bulbar function, which may

indicate early signs of disease progression. The paper presents results from a pilot study demonstrating the feasibility and potential of the proposed system for remote monitoring of ALS patients. The findings suggest that frequent and remote speech analysis can provide valuable insights into the disease course and facilitate timely interventions.

(6) *Zahid, L. et al. (2020) "A Spectrogram-Based Deep Feature Assisted Computer-Aided Diagnostic System for Parkinson's Disease"* [17].

The paper introduces a computer-aided diagnostic system for Parkinson's Disease (PD) based on spectrogram analysis of speech. The authors propose using spectrograms, which visually represent the frequency content of speech signals over time, as input to a deep learning model. They extract deep features from the spectrograms using convolutional neural networks (CNNs) and use these features to classify individuals as either having PD or being healthy controls. The paper presents experimental results demonstrating the effectiveness of the proposed system in detecting PD based on spectrogram analysis. The findings suggest that spectrogram-based deep features can capture subtle speech changes associated with PD, leading to improved diagnostic accuracy.

(7) *Vashkevich, M. and Rushkevich, Y. (2020) "Classification of ALS patients based on acoustic analysis of sustained vowel phonations"* [20].

The study investigates the use of acoustic analysis of sustained vowel phonations for classifying ALS patients. The authors hypothesize that sustained vowel sounds produced by ALS patients exhibit distinct acoustic characteristics compared to healthy individuals. They collect sustained vowel phonations (e.g., "ah", "ee", "oo") from ALS patients and healthy controls, extract acoustic features such as formants, jitter, shimmer, and harmonic-to-noise ratio, and train machine learning models to classify individuals as either ALS or healthy. The paper presents results showing that acoustic analysis of sustained vowel phonations can effectively differentiate between ALS patients and healthy controls, suggesting that these simple speech tasks can provide valuable diagnostic information.

(8) *Kuresan, H. et al. (2021) "Parkinson's disease analysis using speech signal processing"* [15].

The paper provides a comprehensive overview of the use of speech signal processing techniques for analyzing Parkinson's Disease (PD). It explores various speech features, including acoustic and prosodic features, that can be extracted from speech signals to characterize the speech impairments associated with PD. The paper also examines different machine learning algorithms that have been used for PD detection and severity assessment based on speech analysis. It discusses the challenges and opportunities in using speech analysis to improve the diagnosis and management of PD.

(9) *Ding et al. (2024) "Speech based detection of Alzheimer's disease: a survey of AI techniques, datasets and challenges"* [19].

The review paper presents a comprehensive survey exploring the use of AI for detecting Alzheimer's disease from speech. The study aims to provide an overview of the current state-of-the-art, focusing on the different AI techniques employed, the datasets used for training and evaluation, and the challenges faced in this emerging field. Their methodology involves a systematic review and synthesis of existing literature, categorizing and analyzing diverse approaches to speech-based Alzheimer's detection.

Table 9. Chosen studies for deep learning in neurological pathology identification—an analytical overview.

Feature	Cummins et al. (2018) [2]	Vizza et al. (2019) [16]	Ipiña et al. (2020) [18]	Teplansky et al. (2023) [21]	Stegmann et al. (2020) [22]	Zahid et al. (2020) [17]	Vashkevich & Rushkevich (2020) [20]	van Gelderen et al. (2024) [15]	Ding et al. (2024) [19]
Focus	Speech Analysis Overview	Neuro Diseases	MCI	ALS Bulbar Regression	ALS Bulbar Changes (Remote)	PD	ALS	PD	AD
Type of Study	Review	Review	Experimental	Experimental	Experimental	Experimental	Experimental	Review	Review
Target Disease(s)	Multiple	Multiple	MCI	ALS	ALS	PD	ALS	PD	AD
Speech task	Varied	Varied	Elicited	Elicited	Elicited	Elicited	Sustained Vowels	Varied	Spontaneous
Features extracted	Acoustic, Prosodic, Articulatory	Acoustic, Prosodic, Articulatory	Acoustic, Disfluency	Acoustic	Acoustic	Spectrogram-based Deep Features	Acoustic (Formants, Jitter, Shimmer)	From Speech	Acoustic, Lexical, Syntactic, Semantic
Machine learning methods	Varied Overview	Varied Overview	SVM, others	Correlation & Regression	Unspecified	CNN	SVM, others	Varied Overview	Varied Overview
Key findings	DL impact on speech analysis	Methodology for speech analysis	Disfluencies as MCI marker	Speech reflects bulbar function	Remote monitoring feasibility	Spectrograms improve PD detection	Sustained vowels for ALS classification	DL impact on speech analysis	Spontaneous speech for AD detection
Data acquisition	Varied	Varied	Recorded	Recorded	Remote Recording	Recorded	Recorded	Varied	Varied
Study design	Literature Review	Methodologic Review	Case–Control	Case–Control	Longitudinal Pilot	Case–Control	Case–Control	Systematic Review	Literature Review

Table 9 reveals several key characteristics of the included studies. Notably, the studies explore a varied landscape of neurodegenerative diseases, showcasing the wide applicability of speech analysis in this domain. The research approaches are also diverse, encompassing both review papers that offer comprehensive overviews and experimental studies designed to test specific hypotheses. Furthermore, the studies employ a range of speech tasks, including elicited speech such as reading passages and sustained vowels, as well as spontaneous speech, each designed to capture different facets of speech production. In terms of feature extraction, acoustic features are the most prevalent, although some studies also integrate disfluency measures, spectrogram-based features, and even textual features. To analyze these features, a variety of machine learning techniques are utilized for classification and prediction, with deep learning approaches becoming increasingly common in more recent research. Finally, while most experimental studies rely on recorded speech, the study Stegmann et al. from 2020 [22] presents a unique approach, exploring remote speech data acquisition for the purpose of continuous monitoring.

6.2. In-Depth Analysis of the Chosen Studies for This Section

Since Parkinson's Disease is one of the most relevant neurological conditions that affects millions of people worldwide, in this paper's discussion chapter, a decision was made upon choosing one of the more recent and vast experimental studies: Zahid, L. et al. from 2020 "A Spectrogram-Based Deep Feature Assisted Computer-Aided Diagnostic System for Parkinson's Disease" [17] because it is one of the studies with the strongest connection to neural networks.

Study [17] investigates the use of deep learning techniques, specifically leveraging spectrogram images of speech signals, to assist in the diagnosis of Parkinson's Disease (PD). Instead of directly feeding raw audio data into a model, they transform the audio into spectrograms (visual representations of frequencies over time) and then extract deep features from these images using pre-trained CNNs. These features are subsequently used to train classifiers to distinguish between individuals with and without PD.

For sound data input, the study [17] relies on data acquired from the "UCI Parkinson's Disease Database", a collection of voice recordings from both individuals diagnosed with Parkinson's Disease (PD) and healthy controls. Specifically, the analysis in the paper [17] incorporates 195 recordings from PD patients and 63 recordings from healthy individuals. However, it is important to recognize that the dataset possesses limitations due to its lack of diversity, as the subjects are primarily sourced from a single geographic area.

For preprocessing and spectrogram generation, established techniques were employed. While the paper [17] does not explicitly describe audio segmentation, the creation of spectrograms inherently necessitates dividing the continuous audio signal into smaller, overlapping segments using a windowing method. Spectrogram images were then generated by applying the Short-Time Fourier Transform (STFT) to these audio segments, which decomposes the signal into its frequency components over time. The resulting spectrogram visually represents the magnitude of these frequencies, with time on the x-axis, frequency on the y-axis, and intensity indicating the strength of each frequency component. Crucial parameters such as window size, window function type (e.g., Hamming), and Fast Fourier Transform (FFT) length influence the resolution of the spectrogram; however, the specific parameters selected for this study were not detailed. Finally, these spectrograms, represented as matrices of frequency and time values, were converted into image formats like PNG or JPEG to serve as input for the convolutional neural networks (CNNs).

Regarding deep feature extraction, the author's primary approach revolves around leveraging pre-trained convolutional neural networks (CNNs), specifically those trained on extensive image datasets like "ImageNet", as feature extractors. This utilizes the principle

of transfer learning, assuming that these pre-trained networks have acquired generic image features applicable to the task of Parkinson's Disease (PD) diagnosis. Different CNN architectures, such as the classic VGG16, VGG19, ResNet50, InceptionV3, and MobileNetV2, are considered in their experiments. To implement this, the final classification layer of the pre-trained CNN is removed, and the output from an intermediate layer, often a pooling layer or the last convolutional layer, is extracted and used as the feature vector. This vector effectively represents a high-level, compressed representation of the spectrogram image. Subsequently, the output feature maps from the selected layer are flattened into a one-dimensional feature vector, whose dimensionality is determined by both the CNN architecture and the specific layer chosen for extraction.

In the realm of classification, the author [17] concentrates on several key aspects. These include classifier training, where they utilize extracted deep features as input to train a variety of machine learning classifiers, such as Support Vector Machines (SVM), k-Nearest Neighbors (k-NNs), Decision Trees, Random Forests, and "Multi-Layer Perceptrons" (MLPs) [17]. A crucial step involves splitting the dataset into training and testing sets, typically using a 70/30 or 80/20 split, to assess the system's ability to generalize to new, unseen data. To further ensure robust performance evaluation, cross-validation techniques, like "k-fold cross-validation", are employed. Finally, the author [17] addresses hyperparameter tuning, optimizing the performance of the classifiers by adjusting parameters such as the SVM kernel or the number of neighbors in k-NN, often using methods like "grid search" or "randomized search".

The author [17] assesses the performance of the model by using standard classification metrics. Accuracy measures the overall percentage of instances that were correctly classified. Sensitivity, also known as recall or the True Positive Rate, indicates the model's ability to correctly identify patients with Parkinson's Disease. Specificity, or the True Negative Rate, measures the model's ability to correctly identify healthy individuals. Precision represents the proportion of correctly identified PD patients out of all instances predicted to have PD. The F1-Score provides the harmonic mean of precision and recall, offering a balanced measure of performance. Finally, the Area Under the ROC Curve (AUC) quantifies the classifier's ability to distinguish between the two classes across a range of threshold settings.

Because the research paper [17] utilizes a variety of CNN architectures, it may be difficult to provide a definitive conclusion

A summary of possible results using different pre-trained models with different classifiers, is shown in Table 10.

It is considered important to highlight that the study emphasizes some general trends that would be considered useful also for other future research work. Pre-trained CNNs, such as VGG16, ResNet50, and InceptionV3, generally exhibit strong performance due to their capacity for extracting robust features. When used in conjunction with deep features, Support Vector Machines and Random Forests frequently achieve high accuracy, although other classifiers like "k-Nearest Neighbors" (k-NNs) and "Multilayer Perceptrons" (MLPs) also perform well. The selection of the feature extraction layer within the CNN is important and can substantially influence performance, with deeper layers often capturing more task-specific features. Finally, the study's findings may be affected by the limitations of the dataset, which lacks significant diversity.

In summary, the paper [17] presents a computer-aided diagnosis system for Parkinson's Disease using spectrograms and pre-trained CNNs. The method involves converting speech signals into spectrograms, extracting deep features using CNNs, and training classifiers using these features. The study demonstrates the effectiveness of this approach in distinguishing between individuals with and without Parkinson's Disease, achieving promising results in terms of accuracy, sensitivity, specificity, and AUC. However, it is

important to consider the limitations of the dataset used and the potential for further improvements by exploring different CNN architectures, feature selection techniques, and classifier combinations.

Table 10. Results of using different pre-trained models with different classifiers in the 2020 paper by Zahid, L. et al. [17].

Metric	Description	Value Range
Accuracy	Proportion of correctly classified instances.	0–1 (or 0–100%)
Sensitivity	Ability to correctly identify PD patients.	0–1 (or 0–100%)
Specificity	Ability to correctly identify healthy controls.	0–1 (or 0–100%)
Precision	Proportion of correctly identified PD patients among those predicted as PD.	0–1 (or 0–100%)
F1-Score	Harmonic mean of precision and recall.	0–1
Area Under the ROC Curve (AUC)	Area under the Receiver Operating Characteristic curve; measures discriminative ability.	0–1

6.3. Other Significant Work and Observations from Section 6

In the 2023 study by Teplansky et al. [21], the authors sought to address the challenges of quantifying speech impairments in Amyotrophic Lateral Sclerosis (ALS), a neurodegenerative disease that progressively affects motor functions, including those involved in articulation. The study utilized a data-driven articulatory consonant distinctiveness space, constructed from acoustic and articulatory features extracted from speech samples. This innovative method involved collecting data from 50 participants to 25 with confirmed ALS diagnoses and 25 age- and gender-matched healthy controls—who were recorded producing a standardized set of consonant–vowel syllables.

By mapping articulatory movements, the researchers could measure the distinctiveness of consonants based on parameters such as tongue and lip position, and overall articulatory precision.

The primary goal was to identify patterns of impairment in ALS patients, potentially informing diagnostic tools and therapeutic interventions. Results indicated that ALS participants demonstrated reduced articulatory distinctiveness across multiple consonant categories, highlighting the disease's impact on fine motor control in speech production.

Statistical analysis comparing the articulatory consonant distinctiveness scores between the ALS group and healthy controls revealed statistically significant differences. This finding underscores the sensitivity of the data-driven acoustic analysis approach employed in capturing subtle speech motor deficits. Specifically, individuals with ALS demonstrated significantly lower mean distinctiveness scores compared to the control group. This impairment was particularly evident for consonants requiring precise or fine motor control for articulation, such as alveolar and labial phonemes, which are known to be frequently compromised in neurodegenerative motor neuron diseases like ALS. These results demonstrate that quantitative measurements derived from the articulatory consonant distinctiveness space not only capture subtle deficits in speech articulation associated with ALS but also show potential as a quantitative biomarker for monitoring disease progression. To illustrate the quantitative outcomes, Table 11 summarizes key results from the paper [21], including mean articulatory distinctiveness scores for selected consonant categories, standard deviations, and comparisons between ALS patients and controls. These scores were derived

from the articulatory space model, with higher values indicating greater distinctiveness, which scores range from 0 to 1, with 1 representing perfect articulatory separation in the data-driven space.

Table 11. Key results from the 2023 study by Teplansky et al. [21].

Consonant Category	Group	Mean Distinctiveness Score	Standard Deviation
Alveolars (e.g.,/t/,/d/)	Controls	0.82	0.09
	ALS Patients	*0.61*	0.14
Labials (e.g.,/p/,/b/)	Controls	0.79	0.11
	ALS Patients	*0.58*	0.16
Velars (e.g.,/k/,/g/)	Controls	0.75	0.1
	ALS Patients	*0.67*	0.12
Overall Average	Controls	0.78	0.1
	ALS Patients	*0.62*	0.14

The integration of deep learning (DL) with sound-based detection methodologies represents a transformative frontier in the early diagnosis and monitoring of neurological disorders. Conditions such as Parkinson's disease manifest subtle, yet discernible, changes in vocalizations, speech patterns, or other bio-acoustic signals long before overt clinical symptoms appear. Traditional signal processing techniques often struggle with the inherent complexity, non-linearity, and high dimensionality of these audio features. However, deep learning models excel at autonomously learning intricate, hierarchical representations directly from raw audio data, circumventing the need for laborious hand-crafted feature engineering. This unparalleled capacity for complex pattern recognition allows for the identification of nuanced acoustic biomarkers, significantly enhancing diagnostic accuracy, facilitating objective disease tracking, and offering a non-invasive, cost-effective avenue for large-scale screening.

Looking ahead, several promising directions will further solidify the role of deep learning in this domain. A critical next step involves the development of larger, more diverse, and standardized audio datasets to improve model generalization across varied populations and recording environments. Furthermore, integrating multi-modal data fusion, combining audio analysis with clinical, imaging, or genetic information, holds immense potential for a more comprehensive and robust diagnostic picture. Future research will also focus on advancing explainable AI (XAI) techniques to provide clinical interpretability for model predictions, fostering trust and facilitating adoption. Ultimately, the goal is to develop highly robust, real-time deployable systems capable of continuous remote monitoring, paving the way for personalized interventions and improving patient outcomes in neurological care.

7. Frontiers and Challenges: Respiratory, Cardiac, and Neurological Perspectives

Due to the fact that this field is still in its early stages, it is a clear and visible aspect that further research is needed to address challenges and limitations.

By researching this topic and summing up the previous chapters, this paper has identified some of the problematic areas that need to be addressed by further research, improvement of current technical solutions, and interdisciplinary collaborations, as shown in Figure 6.

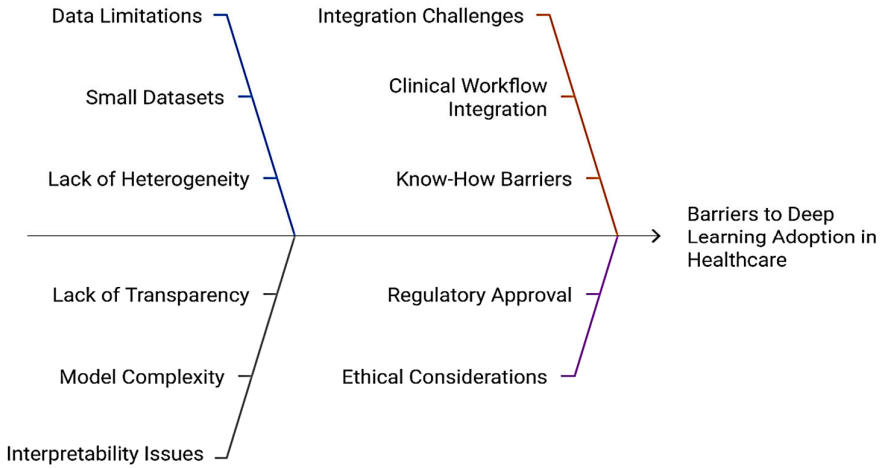

Figure 6. A short summary—challenges when using deep learning for acoustic pathology analysis.

Several challenges, amongst others, impede the wider adoption of deep learning within healthcare. These include the limitations imposed by the typically small, curated datasets used in current research, which potentially fail to represent the diverse nature of real-world patient populations.

Furthermore, the inherent lack of interpretability in many deep learning models poses a significant obstacle, hindering the understanding of the specific acoustic characteristics and patterns driving diagnostic predictions.

Coupled with this is the difficulty in ensuring robust generalization across heterogeneous patient demographics and varied clinical environments, alongside the limited transferability of models for application across different healthcare systems and geographical regions.

A critical requirement is the seamless integration of deep learning solutions into established clinical workflows to facilitate their practical implementation and impact on decision-making.

Concurrently, bridging the knowledge gap between clinicians and technically trained personnel is vital for fostering trust in deep learning-based diagnoses through transparent explanations of underlying reasoning.

Addressing the ethical considerations associated with data privacy, security, and potential algorithmic bias is paramount for responsible deployment.

The often-protracted nature of bureaucratic validation and regulatory approval processes, while necessary to ensure the safety and effectiveness of these tools, can also delay their adoption, necessitating increased engagement from human factors experts and decision-makers.

Finally, the substantial upfront costs and specialized computational resources required for implementation can act as prohibitive factors, restricting the accessibility and scalability of deep learning-based systems in healthcare settings.

8. Bridging the Gap in Deep Learning Solutions for Challenges in Respiratory, Cardiac, and Neurological Pathologies

As deep learning continues to advance and become more integrated into healthcare, several key trends and emerging directions are likely to shape the future of this field.

Table 12 summarizes three major papers that highlight future trends [23–25].

Table 12. The aim of the three papers that look at future research directions in applying deep learning to sound-driven pathology detections.

Study	Aim
1. Azghadi, R. M. et al. (2020) *"Hardware Implementation of Deep Network Accelerators Towards Healthcare and Biomedical Applications"* [25]	To review the existing landscape of hardware accelerators specifically designed for Deep Neural Networks (DNNs) in healthcare and biomedical applications, focusing on their design, performance, and potential benefits.
2. Steiner, F. D., Chen, C. P., and Mermel, H. C. (2020) *"Closing the translation gap: AI applications in digital pathology"* [24]	To examine Table the application of AI, specifically deep learning, in digital pathology, focusing on how it can bridge the gap between research findings and practical clinical implementation. The study also explores the challenges and opportunities for AI adoption in this area.
3. Zhou, K. S. et al. (2021) *"A Review of Deep Learning in Medical Imaging: Imaging Traits, Technology Trends, Case Studies With Progress Highlights, and Future Promises"* [23]	To provide a comprehensive review of deep learning applications in various medical imaging modalities. The study covers imaging traits, technology trends, case studies, progress highlights, and future promises of deep learning in this field.

The proposed papers, while exhibiting a common thematic thread, diverge in their specific focus. Achieving trustworthy, reliable, and clinically validated Artificial Intelligence and deep learning (AI/DL) solutions in healthcare represents a central and unifying objective. Realizing this objective is critically hindered by pervasive challenges, particularly concerning data availability and standardization issues, as well as the fundamental requirement for explainable AI methodologies.

However, the papers differ in their specific areas of emphasis.

In one paper [25], the authors direct their attention specifically towards hardware acceleration strategies, whereas in another paper [24], the authors concentrate their analysis on the application of AI/DL within digital pathology.

In contrast, Zhou et al. [21] present a more comprehensive overview encompassing the broader field of medical imaging. This variation in scope subsequently shapes the targeted future directions advocated by each respective study.

By researching this topic and summing up the previous chapters, this paper has proposed possible solutions that need to be considered by further and ongoing research, in order to efficiently apply deep learning in sound-driven pathology detections, as shown in Figure 7.

Several potential solutions warrant consideration, encompassing but not limited to the ones proposed as follows:

Starting with the curation and dissemination of larger, more diverse datasets that accurately reflect the heterogeneity inherent in real-world clinical environments is paramount.

The development of Explainable Artificial Intelligence (XAI) approaches is essential for enhancing the interpretability of deep learning models, thereby fostering trust and facilitating knowledge transfer between specialists.

Collaborative learning paradigms, such as federated learning and distributed computing frameworks, offer promising avenues for training models on decentralized data while upholding patient privacy and data sovereignty.

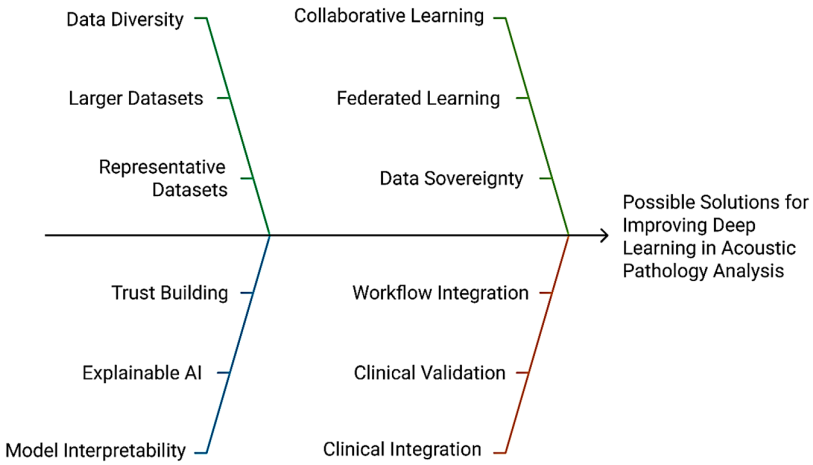

Figure 7. A short summary—solutions to problems that arise from using deep learning for acoustic pathology identification.

Another solution could be continuously evolving neural network architectures, including graph neural networks and transformer models, which represent a fourth area of opportunity for constructing more robust and generalizable models applicable to acoustic pathology analysis.

Advancements in optimization methodologies, model compression techniques, and hardware acceleration are crucial for enabling real-time processing and deployment of deep learning-based tools in clinical settings.

Rigorous clinical validation through large-scale prospective trials, coupled with seamless integration of these tools into existing clinical workflows, is necessary to establish their safety, efficacy, and clinical utility [26,27].

The adoption of multimodal approaches that integrate diverse data sources, such as audio, video, and medical imaging, can provide a more holistic understanding of pathological conditions.

Bridging the gap between technical research and clinical practice is vital to ensure effective integration of deep learning models into clinical workflows, thereby delivering tangible benefits to both healthcare professionals and patients [28].

Also, an engineering approach driven by medical domain knowledge, specifically leveraging the expertise of medical professionals, audiologists, and signal processing specialists, will be crucial for developing clinically relevant and trustworthy deep learning models for pathology detection.

9. Conclusions

The integration of deep learning methodologies within the domain of sound-based pathology detection represents a burgeoning field with profound implications for fundamentally transforming healthcare paradigms. By harnessing the formidable analytical capabilities of deep learning algorithms to process intricate bio-acoustics data, researchers and clinical practitioners are poised to develop highly sophisticated diagnostic and monitoring instruments. These advanced tools enable the precocious identification, continuous surveillance, and effective management of a diverse spectrum of physiological afflictions, particularly those impacting the respiratory and cardiovascular systems, along with other critical bodily functions. But, for more complex disorders like pleural effusions, lung cancer, and pulmonary edema [29], deep learning might be more of a help once those pathologies have already been identified, for example, the field in medical diagnostic imagining, rather than sound-based pathology detection [30].

To unlock the full promise of this rapidly advancing field, it is imperative to surmount several significant hurdles. These include enhancing model interpretability and transparency, mitigating issues stemming from inherent data scarcity, and effectively incorporating nuanced expert domain knowledge into algorithmic design. Crucially, fostering robust interdisciplinary collaboration among machine learning specialists, signal processing engineers, and clinical medical professionals will be paramount for facilitating these advancements. Such concerted efforts are essential to ensure the clinical utility, real-world applicability, and tangible impact of deep learning models employed for sound-based pathology detection.

Realizing the transformative potential of deep learning within the broader healthcare ecosystem is predicated upon the development of models that embody enhanced robustness, heightened interpretability, and seamless clinical integration. These foundational advancements are indispensable for fostering marked amelioration in patient outcomes and are designed to fundamentally reshape the operational landscape of healthcare delivery, moving towards more preventative and personalized care models.

This review paper posits a future where the synergistic convergence of expertise from both cutting-edge researchers and experienced clinicians facilitates the genesis of deep learning systems. These systems will be meticulously optimized for maximal clinical utility, inherently capable of generating actionable, transparently explainable insights, and meticulously engineered for seamless, intuitive integration into existing clinical workflows.

This collaborative paradigm is indispensable for guaranteeing that state-of-the-art artificial intelligence methodologies genuinely yield tangible, quantifiable advantages for individual patients and the broader healthcare ecosystem alike, thereby bridging the gap between technological innovation and practical clinical benefit.

Author Contributions: All authors contributed equally to the conceptualization, methodology, investigation, and writing of this manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research did not receive grants from any funding agency in the public, commercial, or not-for-profit sectors.

Institutional Review Board Statement: Not applicable. This article is a review study involving the synthesis and analysis of previously published literature. As such, it did not involve any primary data collection from human or animal subjects, and therefore, did not require ethical review or approval from an Institutional Review Board.

Informed Consent Statement: This review article is based solely on publicly available, previously published data and research findings. No new primary data was collected for this study. Therefore, no direct interaction with human subjects or animals occurred, and no informed consent was required. The authors have made every effort to accurately and comprehensively represent the existing body of knowledge on the topic, and to avoid misrepresentation or selective reporting of data. All sources are appropriately cited, and due credit is given to the original authors of the research. The authors confirm that no personal data or personal identifiable information was accessed, used, or disclosed in the preparation of this manuscript. This review complies with all relevant legal and regulatory requirements pertaining to data privacy, copyright, and intellectual property. The authors have strived to ensure objectivity and impartiality in the interpretation and synthesis of reviewed literature.

Data Availability Statement: This study is a review article/meta-analysis and did not involve the collection or analysis of original primary data. All information used was obtained from publicly available sources, as cited in the manuscript.

Conflicts of Interest: The authors declare that they have no conflicts of interest.

References

1. Gupta, R.; Chaspari, T.; Kim, J.; Kumar, N.; Bone, D.; Narayanan, S. Pathological speech processing: State-of-the-art, current challenges, and future directions. In Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 20–25 March 2016; pp. 6470–6474.
2. Cummins, N.; Baird, A.; Schuller, B.W. Speech analysis for health: Current state-of-the-art and the increasing impact of deep learning. *Methods* **2018**, *151*, 41–54. [CrossRef]
3. Mehrish, A.; Majumder, N.; Bharadwaj, R.; Mihalcea, R.; Poria, S. A review of deep learning techniques for speech processing. *Inf. Fusion* **2023**, *99*, 101869. [CrossRef]
4. Pham, L.D.; Phan, H.; Palaniappan, R.; Mertins, A.; McLoughlin, I. CNN-MoE Based Framework for Classification of Respiratory Anomalies and Lung Disease Detection. *IEEE J. Biomed. Heal. Inform.* **2021**, *25*, 2938–2947. [CrossRef] [PubMed]
5. Ren, Z.; Chang, Y.; Nguyen, T.T.; Tan, Y.; Qian, K.; Schuller, B.W. A Comprehensive Survey on Heart Sound Analysis in the Deep Learning Era. *IEEE Comput. Intell. Mag.* **2024**, *19*, 42–57. [CrossRef]
6. 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.* **2020**, *25*, 2595–2603. [CrossRef] [PubMed]
7. Pereira, C.A.C.; Soares, M.R.; Boaventura, R.; Castro, M.D.C.; Gomes, P.S.; Gimenez, A.; Fukuda, C.; Cerezoli, M.; Missrie, I. Squawks in interstitial lung disease prevalence and causes in a cohort of one thousand patients. *Medicine* **2019**, *98*, e16419. [CrossRef]
8. Paciej, R.; Vyshedskiy, A.; Bana, D.; Murphy, R. Squawks in pneumonia. *Thorax* **2004**, *59*, 177–178. [CrossRef]
9. Pasterkamp, H.; Kraman, S.S.; Wodicka, G.R. Respiratory Sounds. *Am. J. Respir. Crit. Care Med.* **1997**, *156*, 974–987. [CrossRef]
10. 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.
11. 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.
12. Reed, T.R.; Reed, N.E.; Fritzson, P. Heart sound analysis for symptom detection and computer-aided diagnosis. *Simul. Model. Pr. Theory* **2004**, *12*, 129–146. [CrossRef]
13. Sfayyih, A.H.; Sabry, A.H.; Jameel, S.M.; Sulaiman, N.; Raafat, S.M.; Humaidi, A.J.; Al Kubaiaisi, Y.M. Acoustic-Based Deep Learning Architectures for Lung Disease Diagnosis: A Comprehensive Overview. *Diagnostics* **2023**, *13*, 1748. [CrossRef]
14. Chorba, J.S.; Shapiro, A.M.; Le, L.; Maidens, J.; Prince, J.; Pham, S.; Kanzawa, M.M.; Barbosa, D.N.; Currie, C.; Brooks, C.; et al. Deep Learning Algorithm for Automated Cardiac Murmur Detection via a Digital Stethoscope Platform. *J. Am. Hear. Assoc.* **2021**, *10*, e019905. [CrossRef]
15. Van Gelderen, L.; Tejedor-García, C. Innovative Speech-Based Deep Learning Approaches for Parkinson's Disease Classification: A Systematic Review. *Appl. Sci.* **2024**, *14*, 7873. [CrossRef]
16. Vizza, P.; Tradigo, G.; Mirarchi, D.; Bossio, R.B.; Lombardo, N.; Arabia, G.; Quattrone, A.; Veltri, P. Methodologies of speech analysis for neurodegenerative diseases evaluation. *Int. J. Med Inform.* **2019**, *122*, 45–54. [CrossRef] [PubMed]
17. Zahid, L.; Maqsood, M.; Durrani, M.Y.; Bakhtyar, M.; Baber, J.; Jamal, H.; Mehmood, I.; Song, O.-Y. A Spectrogram-Based Deep Feature Assisted Computer-Aided Diagnostic System for Parkinson's Disease. *IEEE Access* **2020**, *8*, 35482–35495. [CrossRef]
18. López-De-Ipiña, K.; Martinez-De-Lizarduy, U.; Calvo, P.M.; Beitia, B.; García-Melero, J.; Fernández, E.; Ecay-Torres, M.; Faundez-Zanuy, M.; Sanz, P. On the analysis of speech and disfluencies for automatic detection of Mild Cognitive Impairment. *Neural Comput. Appl.* **2018**, *32*, 15761–15769. [CrossRef]
19. Ding, K.; Chetty, M.; Hoshyar, A.N.; Bhattacharya, T.; Klein, B. Speech based detection of Alzheimer's disease: A survey of AI techniques, datasets and challenges. *Artif. Intell. Rev.* **2024**, *57*, 325. [CrossRef]
20. Vashkevich, M.; Rushkevich, Y. Classification of ALS patients based on acoustic analysis of sustained vowel phonations. *Biomed. Signal Process. Control.* **2021**, *65*, 102350. [CrossRef]
21. Teplansky, K.J.; Wisler, A.; Green, J.R.; Heitzman, D.; Austin, S.; Wang, J. Measuring Articulatory Patterns in Amyotrophic Lateral Sclerosis Using a Data-Driven Articulatory Consonant Distinctiveness Space Approach. *J. Speech Lang. Hear. Res.* **2023**, *66*, 3076–3088. [CrossRef]
22. Stegmann, G.M.; Hahn, S.; Liss, J.; Shefner, J.; Rutkove, S.; Shelton, K.; Duncan, C.J.; Berisha, V. Early detection and tracking of bulbar changes in ALS via frequent and remote speech analysis. *npj Digit. Med.* **2020**, *3*, 132. [CrossRef]
23. Zhou, S.K.; Greenspan, H.; Davatzikos, C.; Duncan, J.S.; Van Ginneken, B.; Madabhushi, A.; Prince, J.L.; Rueckert, D.; Summers, R.M. A Review of Deep Learning in Medical Imaging: Imaging Traits, Technology Trends, Case Studies With Progress Highlights, and Future Promises. *Proc. IEEE* **2021**, *109*, 820–838. [CrossRef]
24. Steiner, D.F.; Chen, P.-H.C.; Mermel, C.H. Closing the translation gap: AI applications in digital pathology. *Biochim. Biophys. Acta (BBA)-Rev. Cancer* **2021**, *1875*, 188452. [CrossRef]

25. Azghadi, M.R.; Lammie, C.; Eshraghian, J.K.; Payvand, M.; Donati, E.; Linares-Barranco, B.; Indiveri, G. Hardware Implementation of Deep Network Accelerators Towards Healthcare and Biomedical Applications. *IEEE Trans. Biomed. Circuits Syst.* **2020**, *14*, 1138–1159. [CrossRef]
26. Serag, A.; Ion-Margineanu, A.; Qureshi, H.; McMillan, R.; Saint Martin, M.J.; Diamond, J.; O'Reilly, P.; Hamilton, P. Translational AI and Deep Learning in Diagnostic Pathology. *Front. Med.* **2019**, *6*, 185. [CrossRef] [PubMed]
27. Misu, R.S. Brain Tmor Detection Using Deep Learning Approaches. *arXiv* **2023**, arXiv:2309.12193. [CrossRef]
28. Adlung, L.; Cohen, Y.; Mor, U.; Elinav, E. Machine learning in clinical decision making. *Med* **2021**, *2*, 642–665. [CrossRef] [PubMed]
29. 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]
30. Ardila, D.; Kiraly, A.P.; Bharadwaj, S.; Choi, B.; Reicher, J.J.; Peng, L.; Tse, D.; Etemadi, M.; Ye, W.; Corrado, G.; et al. Author Correction: End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat. Med.* **2019**, *25*, 1319. [CrossRef]