


A Comprehensive Survey on Heart Sound Analysis in the Deep Learning Era

Zhao Ren 

*Leibniz University Hannover, GERMANY and also
University of Bremen, GERMANY*

Yi Chang 


Imperial College London, U.K.

Thanh Tam Nguyen 

Griffith University, AUSTRALIA

Yang Tan  and **Kun Qian** 

Beijing Institute of Technology, CHINA

Björn W. Schuller 

*University of Augsburg, GERMANY, also Technical
University of Munich, GERMANY, and also Imperial
College London, U.K.*

Abstract—Heart sound auscultation has been applied in clinical usage for early screening of cardiovascular diseases. Due to the high demand for auscultation expertise, automatic auscultation can help with auxiliary diagnosis and reduce the burden of training professional clinicians. Nevertheless, there is a limit to classic machine learning’s performance improvement in the era of Big Data. Deep learning has outperformed classic machine learning in many research fields, as it employs more complex model architectures with a stronger capability of extracting effective representations. Moreover, it has been successfully applied to heart sound analysis in the past years. As most review works about heart sound analysis were carried out before 2017, the present survey is the first to work on a comprehensive overview to summarise papers on heart sound analysis with deep learning published in 2017–2022. This work introduces both classic machine learning and deep learning for comparison, and further offer insights about the advances and future research directions in deep learning for heart sound analysis.

This article was recommended for publication by Associate Editor Ah-Hwee Tan. Corresponding author: Zhao Ren (e-mail: zren@uni-bremen.de).

I. Introduction

Cardiac auscultation, i.e., listening to and interpreting the heart sound, is an indispensable and critical part of the clinical examination of the patient [1]. As a low-cost and non-invasive examination, cardiac auscultation is invaluable for detecting a heart disease and providing an estimate of its severity, evolution, and prognosis [2]. Accurate cardiac auscultation may determine whether a more expensive and throughout examination should be conducted [2]. Nevertheless, due to difficulties in diagnosing diastolic murmurs, the overall sensitivity of cardiac auscultation is poor (i.e., ranging from 0.21 to 1.00) [1]. Poor cardiac auscultation skills may either overlook significant pathology, causing deteriorating condition, or overdiagnose pathology, leading to inappropriate referral for expensive echocardiography [1].

To tackle the above problems of cardiac auscultation, classic machine learning (ML) has been widely used for automated heart sound analysis, including denoising, segmentation, and classification. For instance, support vector machines (SVMs) have been employed to detect noisy audio clips [3], and hidden Markov models (HMMs) have been used for heart sound segmentation [4]. Classifiers like SVMs and decision trees have been applied to heart sound classification [5], [6]. They often take hand-crafted acoustic features as the input, however, human knowledge is required for manually selecting features. Additionally, classic ML tends to excel in small-scale data, whereas its performance on large-scale datasets has remained a bottleneck.

More recently, deep learning (DL) has demonstrated its higher capability in analysing heart sounds than classic ML [7]. DL models typically accept raw audio signals or time-frequency representations as the inputs [8], [9], thereby improving the efficiency by skipping the need for selecting hand-crafted acoustic features. Complex structures of DL models also enhance their ability to learn abstract representations from large-scale datasets.

A. Differences Between This Survey and its Precursors

There are several review studies on heart sound analysis (see Table I). Hand-crafted feature extraction and basic ML models (e.g., SVMs and shallow artificial neural networks) were discussed in [10], [11], [12], [13], [14]. DL has been used since 2016 in the PhysioNet challenge [15]. Only the studies in [15], [16] discussed heart sound analysis, including denoising, segmentation, and classification, along with DL approaches for partial analysis tasks. Additionally, the study in [17] explored DL for heart sound classification. Nevertheless, heart sound segmentation with DL was not included in [15], [16], [17]. This survey fills the gap in the existing reviews, as few studies have provided a comprehensive review of DL methods for heart sound analysis since 2017. Furthermore, the state-of-the-art approaches regarding the interpretability of DL models will be summarised and discussed. This work will also summarise multiple heart sound

TABLE I A comparison of existing surveys on heart sound analysis. Den.: Denoising, Seg.: Segmentation, Cla.: Classification, Int.: Interpretation, mo: Mentioned only.

SURVEYS	YEAR	DL	DEN.	SEG.	CLA.	INT.
Bhoi et al. [10]	by 2012	X	mo	✓	✓	X
Chakrabarti et al. [11]	by 2013	X	✓	✓	✓	X
Nabih-Ali et al. [12]	2004-2016	X	✓	✓	✓	X
Clifford et al. [15]	2016-2017	✓	✓	✓	✓	X
Ghosh et al. [13]	by 2018	X	✓	✓	✓	X
Majhi et al. [17]	2008-2018	✓	X	X	✓	X
Dwivedi et al. [14]	1963-2018	X	X	✓	✓	X
Chen et al. [16]	2016-2020	✓	mo	mo	✓	X
This survey	2017-2022	✓	✓	✓	✓	✓

databases, discuss the potential research problems, and outline future research directions to help promote relevant research studies.

B. Challenges in Heart Sound Analysis

Although many ML and DL methods have been applied to heart sound analysis, this research field still faces many technical challenges, including denoising, segmentation, classification, and DL model explanation.

The first challenge is denoising, which aims to remove noise from heart sounds. Recording environments can be noisy with environmental noise and speech, making denoising an essential pre-processing procedure to improve the audio quality for better segmentation and classification performance.

The second challenge is segmentation that splits a heart sound signal into multiple parts, i.e., cardiac cycles or smaller segments (S1, systole, S2, and diastole). Heart sound segmentation is often a pre-processing step for classification.

The third challenge is classification, which predicts the severity level of cardiovascular diseases or heart abnormalities from heart sounds. Heart sound classification is helpful for early screening of heart diseases in primary care.

The final one is explaining DL models for heart sound analysis. DL models, with their complex structures, often appear as black boxes to humans, despite their promising performance in heart sound analysis. In the sensitive domain of digital health, explainable DL models are crucial for clinicians to provide timely and appropriate therapies for patients. Correspondingly, trust from clinicians and patients can promote real-life application of explainable DL models.

C. Contributions of This Survey

The survey has the following contributions.

The first comprehensive survey in heart sound analysis with DL: In addition to summarising ML techniques for heart sound denoising, segmentation, and classification, this work

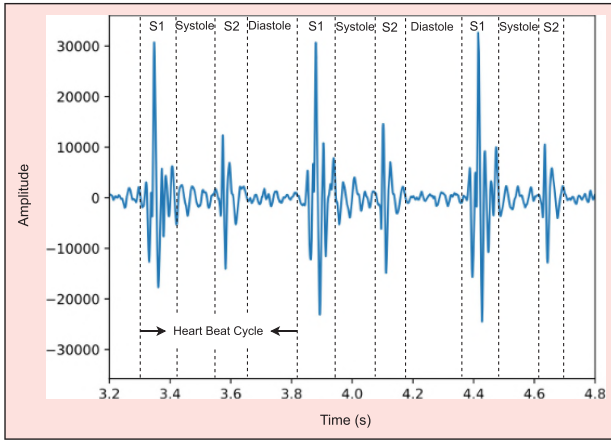


FIGURE 1 The PCG recording of a *normal* heart sound from the PhysioNet/CinC Database [25]. Frames in the middle with four states (i. e., S1, systole, S2, and diastole) are depicted.

reviews advanced DL topologies for heart sound analysis, especially segmentation, classification, and interpretation.

Summarisation of resources: This work summarises publicly available datasets for heart sound analysis, particularly for classification. Additionally, it provides a collection of open-sourced DL algorithms for heart sound classification.

Future research directions: This survey discusses the limitation of current DL methods and points out potential future research topics in this area. It also discusses the importance of explainable DL models for heart sound classification, current advances, and future directions in explainable AI.

II. Background

A. Heart Sounds

In a human's cardiac system, a normal cardiac cycle contains the first heart sound S1 and the second heart sound S2. S1 occurs with the closure of the mitral and tricuspid valves at the start of the systole phase, while S2 is caused by the closure of the aortic and pulmonary valves between the systole and diastole phases [18] (See Figure 1). Additionally, extra heart sounds, i.e., the third heart sound S3 and the fourth heart sound S4, can occur in both normal and pathological conditions [19]. Both S3 and S4 manifest during the diastole phase. Specifically, S3 occurs after S2, resulting from the rapid filling of the ventricles, while S4 occurs before S1 (i.e., at the end of diastole) during the ventricle's late filling due to atrial contraction [8], [18], [20]. The frequency range of S1 and S2 is 20–200 Hz, while that of S3 and S4 ranges between 15–65 Hz [21]. The two types of extra sounds, S3 and S4, may indicate diseases: S3 could be a sign of heart failure [22], and a pathologic S4 is commonly caused by conditions that can result in ventricular hypertrophy [20].

Additionally, a murmur could indicate defective valves or an orifice in the septal wall [22]. Murmurs, caused by

turbulent blood flow in the heart system, are identified as abnormal sounds, and are crucial for diagnosing cardiovascular diseases [23]. Murmurs often constitute the primary basis for diagnosing valvular heart disease [24]. Clinically, murmurs consist of two types: systolic murmurs and diastolic murmurs. Aortic stenosis, mitral regurgitation, and tricuspid regurgitation occur during systole, while mitral stenosis and tricuspid stenosis occur during diastole [23].

B. Diagnosis of Cardiovascular Diseases

Nowadays, several non-invasive diagnostic tools are available for cardiovascular diseases. Particularly, medical imaging tools are capable of visualising the cardiovascular system. For instance, the echocardiogram (echo) utilises ultrasound scans to create a moving picture of the heart, offering insights into its size, shape, structure, and function [26]. Cardiac computed tomography (CT) uses X-rays to create detailed images of the heart and its blood vessels [26]. For assessment of the cardiovascular system's function and structure, cardiac magnetic resonance imaging (CMRI) creates both still and moving pictures of the heart and major blood vessels [26]. However, these imaging instruments are expensive and require trained medical professionals for operation, limiting their application in clinics and small- to medium-sized hospitals.

Compared to the aforementioned diagnostic instruments, cardiac auscultation is low-cost and essential in preliminary physical examinations. Phonocardiogram (PCG) signals, recorded with a phonocardiograph, have proven to be valuable in pediatric cardiology, adult cardiology, and internal medicine [27]. Recent advances in electronic stethoscopes facilitated computer-aided auscultation by integrating sensor design, signal processing, and ML techniques [27]. The low-cost and portable nature of electronic stethoscopes makes it feasible to apply computer-aided auscultation to primary care and remote/home healthcare settings.

Apart from PCG, Electrocardiogram (ECG), which senses the P-QRS-T wave to depict the electrical activity of the heart [28], is an inexpensive and commonly used tool for screening heart diseases. ECG and PCG are highly interrelated as they are concurrent phenomena during heart activities [29]. In an ECG signal, the P wave represents activation of the atria, followed by QRS complex resulting from ventricular excitation [29]. The ventricles then relax back to the electrical resting state, and the T wave shows the ventricular repolarization [29]. During this procedure, S1 occurs when the ventricles contract and the atrioventricular valves close; S2 happens when the ventricles relax and semilunar valves close [29]. Both ECG and PCG have been used for heart abnormality detection [30], [31]. In [32], ML models with both ECG and PCG as inputs outperformed models with only one type of signal for heart abnormality detection. Compared with PCG, ECG has difficulty in detecting structural abnormalities in heart valves and defects characterised by heart murmurs [33]. In

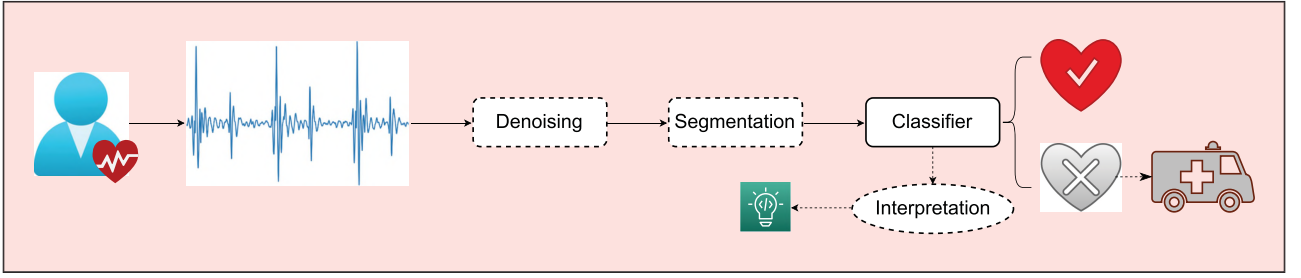


FIGURE 2 The framework of heart sound analysis involves denoising and segmentation, followed by training a classifier to produce the predictions and interpretations for clinicians and patients. The ✓ indicates a normal prediction, while the × is an abnormal one. The dashes “- -” denote optional procedures.

this context, analysing PCG is complementary to ECG analysis in diagnosis.

In Figure 2, heart sounds are processed by denoising, segmentation, and classification, and then clinicians and patients receive the predictions and interpretations in primary care. In real life, patients with heart sounds predicted as abnormal are recommended to do further professional medical examinations for accurate diagnosis.

III. Heart Sound Analysis Tasks

This section describes the tasks and summarises classic ML techniques for each problem, as shown in Figure 3.

A. Denoising

Generally, recorded heart sounds consist of many kinds of noises [34], including white noise and other sounds

presented in the recording environments, e.g., human speech. Noise may impair the segmentation and classification performance of heart sounds [34]. In this regard, numerous studies have explored denoising methods for better performance in heart sound segmentation and classification tasks.

Filters: As a preprocessing procedure of heart sound classification, many denoising approaches employed signal filters to remove noise from heart sounds [8]. Highpass filters have been used to eliminate low-frequency noise [35], [36]. With the capability of mitigating both high- and low-frequency noises, bandpass filters are more often used for heart sound denoising [37]. Butterworth bandpass filters have been successfully employed in many studies [38], [39]. The cutoff frequency of a Butterworth bandpass filter is set with a low frequency for filtering out noise with very low frequencies and a high

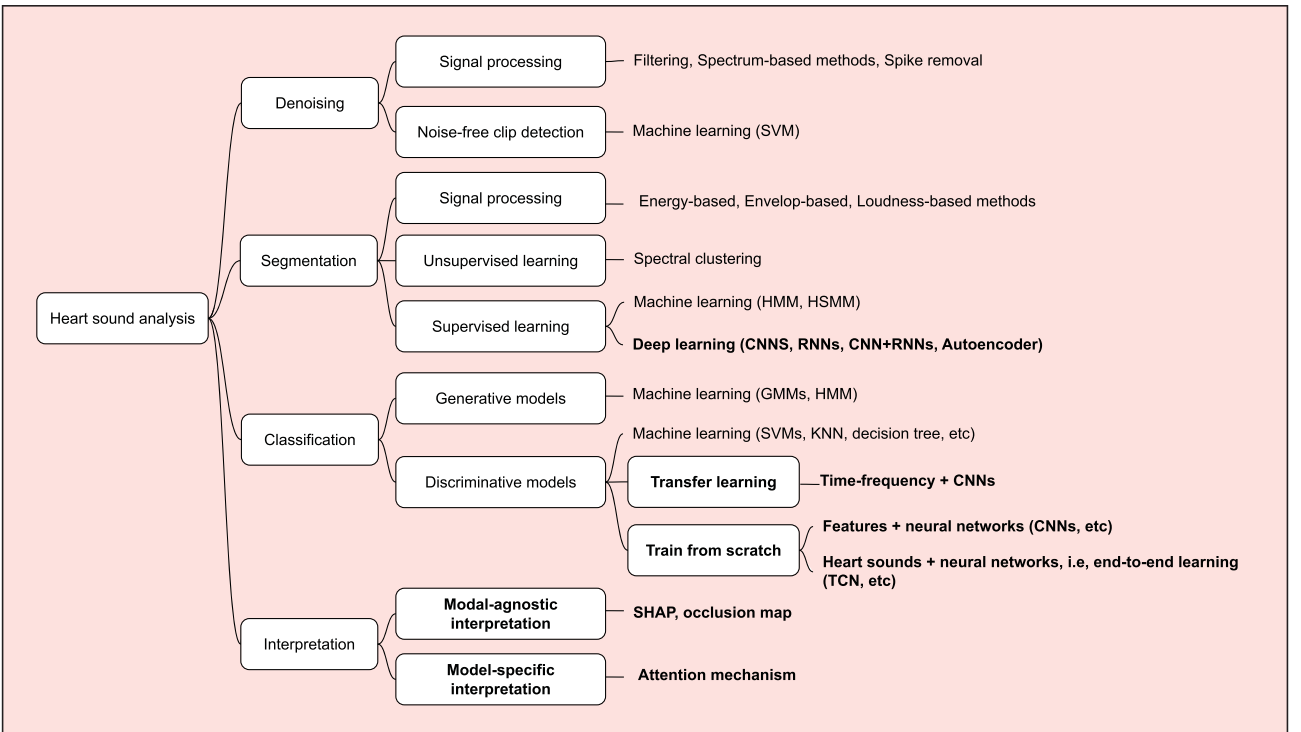


FIGURE 3 Categorisation of methods for heart sound analysis. Bold texts are DL approaches.

frequency for filtering out high-frequency noises. A range of Butterworth bandpass filters with various orders have been applied with different cutoff frequency settings. For instance, a 4-th order Butterworth filter was set with a cutoff frequency of 25–400 Hz in [40], and a 5-th order Butterworth filter was designed to have a cutoff frequency of 25–500 Hz in [41] and 25–250 Hz in [42]. A 6-th order Butterworth filter was designed with a cutoff frequency of 50–950 Hz [43] and 30–900 Hz in [44]. Additionally, several other filters were also used for denoising heart sounds, such as Savitzky-Golay filter [45], Chebyshev low-pass filter [46], and Notch filter [47].

Spectrum-based denoising: To remove noise, spectrograms were simply selected with a threshold of -30, -45, -60, or -75 dB in [48]. However, it is time-consuming to search for a suitable threshold among different heart sounds. A more flexible method, spectral subtraction [49], was used to estimate the noise and remove it from heart sounds [39].

Spike removal: Frictional spike is a redundant part of the amplitude of a heart sound. In several studies [38], [40], frictional spikes have been detected and eliminated (i.e., replaced by zeros) during pre-processing of heart sounds.

Selection of noise-free segments: Apart from removing noise from heart sounds, the usage of noise-free heart sound segments has been regarded helpful for heart sound analysis. Wavelet entropy was used as a feature to evaluate noise in heart sound segments [6], as clean heart sounds have relatively higher wavelet entropy than noisy heart sounds. Empirical wavelet transform was used to separate heart sounds, murmurs, low-frequency artifacts, and high-frequency noises in another study [19]. Additionally, classic ML was also used for detecting noise-free heart sound segments. In [3], SVMs were applied to classify the quality of heart sound signals into binary classes (i.e., ‘unacceptable’ and ‘acceptable’) or three classes (i.e., ‘unacceptable’, ‘good’, ‘excellent’).

Among the above denoising methods, filters are often used to filter out noise with defined frequency bands. Filtering, as a basic denoising approach, can be combined with other denoising methods to further improve the audio quality [39]. Spectral subtraction, which estimates the noise power from the frequencies outside the heart sound frequency range [39], is more flexible than filters. Differently, spike removal focuses on removing the spikes rather than removing the audio components with specific frequency bands. The selection of noise-free segments is more complex than the signal-processing-based methods mentioned above, but it is more effective in automatically selecting audio segments with acceptable qualities before applying other potential denoising methods.

B. Segmentation

Heart sound segmentation aims to split an audio sample into a set of smaller audio segments, which could be equal to or shorter than a complete cardiac cycle [50]. The segments shorter than a cardiac cycle could include S1, systole, S2, and diastole, as indicated in Figure 1.

Energy-based segmentation: As heart sounds at different states have various energies, signal energy has been used for localising S1 and S2 peaks [51]. Based on frequency information (e.g., Wavelet Transform (WT)) of heart sounds, energy peaks of wavelet coefficients were detected for localising S1 and S2 in [52].

Envelope-based segmentation: Apart from energy, heart sound segmentation can be achieved based on envelopes [42], [53]. For instance, Shannon energy envelope was extracted for heart sound segmentation in [54]. In [8], heart sound segmentation was implemented based on the Shannon energy envelope and zero crossings of heart sounds. In [43], [55], S1 of the first heart cycle was detected based on Shannon energy envelopes, and subsequent S1 heart sounds were detected using a sliding heart cycle window.

Loudness-based segmentation: Loudness has proven its potential to segment heart sounds [56], [57]. Specifically, spectrograms extracted from heart sounds are firstly converted into the Bark scale and smoothed using a Hanning window. At each time frame, the sensation of loudness is then calculated by the mean of the amplitudes at all frequency bands: $L(t) = \frac{\sum_{i=1}^T A(i)}{T}$, where $A(t)$ is the amplitude at the t -th time frame, and T is the total number of time frames. Furthermore, the derivation of the loudness function is computed to obtain peaks. Therefore, systoles and diastoles can be localised as they have different time lengths.

Classic Machine Learning for Segmentation: ML models have been proposed for heart sound segmentation, aiming to achieve more noise-robust results than the rule-based segmentation methods mentioned above [58].

Unsupervised Learning: Considering the limited availability of the heart sound datasets, the authors of [59] adopted an unsupervised spectral clustering technique based on Gaussian kernel similarity to obtain frame labels (e.g., S1 and S2), which are further utilised to segment heart sounds.

Supervised Learning: Hidden Markov models (HMMs) have been widely used for segmentation [4]. Let us assume heart states as $S = \{s_1, s_2, s_3, s_4\} = \{S1, \text{systole}, S2, \text{diastole}\}$, and the observations $O = \{o_1, o_2, \dots, o_T\}$ as raw heart sounds or acoustic features. A transmission matrix $A = \{a_{ij}\}$ denotes the probability of a state s_i at the t -th time frame moving to s_j at the $(t + 1)$ -th time frame. The probability density distribution of o_t to be generated by s_i is $B = b_i(o_t) = P[o_t|s_i]$, where P means probability. The initial state distribution is $\pi = \{\pi_i\}$, representing the probability of state s_i at the starting time frame. With A , B , π , and O , an HMM model aims to optimise the state sequence. The Viterbi algorithm, often used for this purpose [58], is further detailed in [4].

To better capture the abrupt changes in PCG signals, the study in [60] used signal envelopes. The kurtosis of the envelope was then computed to extract impulse-like characteristics. Subsequently, these characteristics were passed through a zero-frequency filter to obtain pure impulse information. Along with the heart sound labels, the extracted features were fed into a hidden semi-Markov

model (HSMM). In [61], a multi-centroid-duration-based HSMM was introduced to better adapt to the variability of heart cycle durations (HCDs) in PCG recordings. HCDs were estimated at various instances of a PCG to obtain maximum possible duration values, and those nearest values were clubbed into clusters to refer to each centroid. With more accurate state duration information, the HSMM achieved better segmentation performance. Similarly, considering the inter-patient variability, the emission probability distributions to each patient were estimated through a Gaussian mixture model (GMM), and an improved HSMM was used for segmentation in an unsupervised and adaptive way [62]. Moreover, the expectation maximisation algorithm developed in [63] searched for sojourn time distribution parameters of an HSMM for each subject. Many studies [64], [65], [66] have employed logistic regression-based HSMM (LR-HSMM) [67] for heart sound segmentation. LR was incorporated to predict the probability of $P[s_j|o_t]$, and B was then computed with Bayes' rule. There are also other improved HMM methods, such as the *duration-dependent HMM* [58], [68] considering the probability density function of the duration at each state. Another study [38] proposed a Markov-switching model for heart sound segmentation.

The energy-based, envelop-based, and loudness-based segmentation approaches attempt to detect the corresponding feature peaks, indicating S1 and S2 heart sounds. However, these approaches are primarily applied to high-quality audio samples after the denoising procedure. Classic ML methods have proven more efficient and precise in segmenting noise-contaminated heart sounds even without denoising [58], [59], [67]. One can select unsupervised or supervised learning based on whether segmentation-related labels are available.

C. Classification

The goal of automated auscultation is heart sound classification, including i) detecting abnormal heart sounds (e.g., murmurs, mitral stenosis, etc.) and ii) recognising severity of cardiovascular diseases (normal/mild/moderate).

Feature Engineering: Feature extraction is often performed before training a classifier. Low-level descriptors (LLDs) and functionals are typically extracted as acoustic features. LLDs represent segmental features obtained from short-time segment analysis (see Table II), while functionals are supra-segmental feature vectors derived from LLDs. Functionals generally refer to statistical features such as mean, maximum, standard deviation, and others.

The LLDs used for heart sound classification are listed in Table II. In addition to time-domain LLDs, frequency-domain LLDs have been widely used for heart sound classification. Frequency-domain LLDs include (Mel-scaled) spectral features and wavelet features. There are also existing feature sets used for heart sound classification, such as the COMPARE feature set [69] and the eGeMAPS feature set [70].

In addition to hand-crafted features, more recent studies [71], [72] have explored deep representations (see Table II). Given the strong capabilities of DL models in extracting abstract features, deep representations have the potential to enhance the performance of hand-crafted features.

Classic Machine Learning for Classification: Rule-based methods were proposed for heart sound classification in [19], [94]. To achieve better performance, most research studies have used classic ML for heart sound classification.

Generative models aim to generate the joint probability distribution $P(X, y)$, given the features X and the labels y . The posterior probability $P(y|X)$ is computed via Bayes' rule $P(y|X) = \frac{P(X, y)}{P(X)} = \frac{P(X|y)P(y)}{P(X)}$, where $P(X|y)$ is the likelihood probability distribution. The Naïve Bayes Classifier was widely used in heart sound classification due to its ease of use [18]. Gaussian Mixture Models (GMMs) were used to estimate data distribution by optimising the weights of Gaussian mixture components and the mean and variance in each component [56], [88]. A Gaussian mixture-based HMM [38] was employed for heart sound classification, considering the four sequential heart states. In [89], multiple HMMs without GMMs were used for heart sound classification.

Discriminative models are designed to directly predict the posterior probability $P(y|X)$ given X . Figure 4 presents statistics of recent works from 2017 to 2022 that employ classic ML models for heart sound classification. SVMs have been very widely used by learning a supporting hyperplane between classes [43], [48], [90]. Apart from linear projection between data samples and labels, SVMs can learn to separate hyperplanes on non-linear data via non-linear kernels, such as the radial basis function.

Furthermore, k -nearest neighbours (KNNs) has shown good performance in heart sound classification by classifying a data sample according to the classes of its k -nearest neighbours [73], [80]. Also, decision trees have been successfully employed in [71], [87]. One reason is that limiting the number of decision nodes can help avoid overfitting [6]. Additionally, the structure of a decision tree can reveal the internal logic for classification. Bagged trees assemble multiple decision trees to create more complex model architectures for better performance [71], [75]. Random forests further improve bagged trees with fewer features when splitting each node [80], [92].

In recent years, feed-forward neural networks (FNNs) have been applied to heart sound classification [48], [79]. FNNs can automatically learn a non-linear projection between acoustic features and labels. Despite FNNs' limitation in explainability compared to classifiers like SVMs and decision trees, they show potential to achieve good performance.

There are also several other ML models, such as linear discriminant classifiers [48], [71], logistic regression [95], quadratic discriminant analysis [42], boosting methods [50], and others [47], [74]. Finally, multiple classifiers can be further combined to enhance performance beyond what single models achieve [50], [84].

TABLE II Hand-crafted features and deep representations for heart sound classification.

GROUP	FEATURES	DESCRIPTION	REFERENCE
Time-domain	Envelope	Envelope of a signal	[19], [73]
	Amplitude	Amplitude of a signal	[52]
	Energy	Energy of a signal	[36], [42], [51]
	Entropy	Signal entropy	[36], [74]
	Loudness	Perception of sound magnitude	[56]
	Peak amplitude	Amplitude of peaks	[75]
Spectral	Spectral amplitude	Fourier transform	[6], [36], [38], [42], [74], [76], [77], [78], [79]
	Dominant frequency value	Frequency which leads to the maximum spectrum	[76], [77]
	Dominant frequency ratio	Ratio of the maximum energy to the total energy	[76], [77]
	Energy	Spectral energy	[38]
	Spectral roll-off	Frequency below a percentage of the total spectral energy	[36]
	Spectral centroid	Average of magnitude spectrogram at each frame	[36], [79]
	Spectral flux	Changing speed of the power spectrum	[36]
	Power spectral density (PSD)	Distribution of power in spectral components	[42], [73], [79]
	Spectral entropy	Shannon entropy of PSD	[47], [74], [76], [77]
	Instantaneous frequency	Frequency for non-stationary signals	[80]
	Fractional Fourier transform entropy	Spectral entropy of the fractional Fourier transform	[81]
	Spectrogram	Short-Time Fourier Transform (STFT)	[43], [55]
	Cepstrum	Inverse Fourier transform on the logarithm of the spectrum	[78]
Mel frequency	Mel-frequency	Mel-scaled frequency	[82], [83]
	Mel-Frequency Cepstral Coefficients (MFCCs)	Discrete cosine transform of Mel-scaled spectrogram	[38], [65], [74], [76], [84], [85], [86], [87], [88], [89]
	Fractional Fourier transform-based Mel-frequency	Mel-frequency from the fractional Fourier transform	[39]
Wavelet	Wavelet transform	Frequency analysis of a signal at various scales	[50], [73], [79]
	Wavelet scattering transform	"Wavelet convolution with nonlinear modulus and averaging scaling function" ^a (translation invariance and elastic deformation stability [90])	[90], [91]
	Wavelet synchrosqueezing transform	Reassignment of wavelet coefficients	[92]
	Tunable quality wavelet transform	"Wavelet multiresolution analysis with a user-specified Q-factor, the ratio of the centre frequency to the bandwidth of the filters" ^b	[46], [93]
	Wavelet entropy	Temporal energy distribution based on wavelet coefficients	[6]
Feature set	CoMPaRE	Computational Paralinguistics Challenge feature set	[82]
	eGeMAPS	The extended Geneva Minimalistic Acoustic Parameter Set	[82]
Deep representation	Graph-based features	Petersen graph pattern	[71]
	Sparse coefficient	Result of sparse coding	[5]
	Autoencoder-based features	Features extracted by an autoencoder from hand-crafted features	[48], [72]

^a[Online]. Available: <https://de.mathworks.com/help/wavelet/ug/wavelet-scattering.html>^b[Online]. Available: <https://de.mathworks.com/help/wavelet/ug/tunable-q-factor-wavelet-transform.html>

The introduced hand-crafted features and deep representations can be combined for classification. Compared to hand-crafted features, deep representations have limited explainability, making them unsuitable for interpretable decision trees. Among hand-crafted features, the listed feature groups in Table II (i.e., time-domain, spectral, Mel frequency, and Wavelet features) are complementary in representing heart sound characteristics. Therefore, they are incorporated in the

CoMPaRE and eGeMAPS feature sets. To enhance performance, feature selection methods can be employed to remove redundant/low-contribution features.

Generative models may require more data to model the data distribution, while discriminative models may be more susceptible to outliers. It is observed that more studies have utilised discriminative models for heart sound classification, tending to achieve good performance. However,

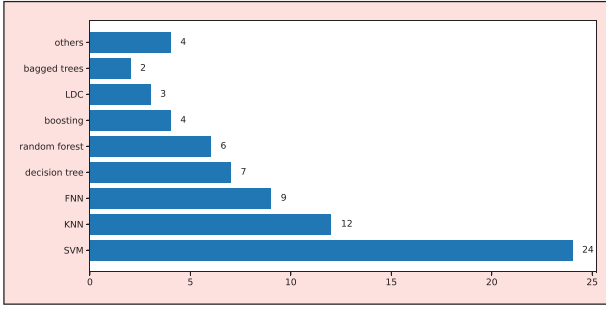


FIGURE 4. Statistics of the literature using discriminative ML models for heart sound classification during 2017–2022. FNN: Feed-forward neural network, KNN: k -nearest neighbour, LDC: Linear discriminant classifier.

generative models can be employed to generate additional data samples based on the learnt data distribution.

IV. State-of-the-Art Deep Learning Studies

DL has been successfully applied to heart sound analysis [7]. As there are few works on heart sound denoising using DL, methods for segmentation and classification with DL are introduced.

A. Deep Learning for Segmentation

Various DL models, categorised into convolutional neural networks (CNNs) for extracting spatial representations and recurrent neural networks (RNNs) for sequential representations, have been proposed for heart sound segmentation.

Convolutional neural networks: Inspired by successful applications of deep CNNs in image segmentation, recent studies have applied deep CNNs to heart sound segmentation [96]. For instance, several CNN-based segmentation algorithms were proposed and compared in [96], including CNNs with sequential max temporal modelling, CNNs with HMMs or HSMMs to model the probability density distribution of observations.

Recurrent neural networks: Given their capacity to leverage temporal information in sequential data, RNNs can aid in identifying the states of heart sounds. In [97], segmentation was approached as an event detection task, leading to the development of bi-directional Gated Recurrent Unit (GRU)-RNNs utilising spectrogram and envelop features. Recognising that envelope features may inadequately capture the intrinsic duration information of heart cycles, a duration-LSTM was proposed in [98]. This model integrated the duration vector into the standard LSTM cells along with envelope features, with the aim of achieving enhanced segmentation performance. Duration parameters encompass heart cycle duration and systole duration estimated from the envelope autocorrelation. In [99], the authors employed bi-directional GRU-RNNs directly for heart sound segmentation, without utilising envelopes and time-frequency based features. Addressing the potential presence of noisy and irregular sequences in heart sound signals, an attention-based RNN framework was introduced in [100]. Specifically, preceding the final classification layer, a single

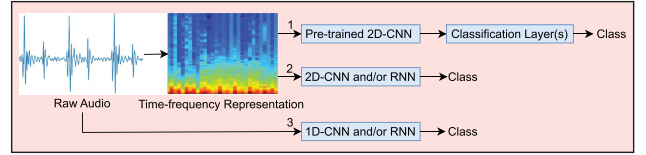


FIGURE 5. Pipeline of DL models working on heart sounds. “1” indicates transfer learning; “2” illustrates deep learning on the time-frequency representation; “3” depicts end-to-end learning. The three branches can either operate in parallel or be assembled at the feature or decision level. Additionally, DL can be utilised for processing features other than raw audio signals and time-frequency representations.

linear layer was applied to the hidden representation returned by bi-directional LSTMs to learn the weight score of each hidden state. These weight score values are then multiplied with the hidden representation to obtain the final classification.

CNNs + RNNs: In [101], an end-to-end model was proposed, integrating CNNs and LSTM recurrent neural networks (RNNs) to directly learn rich and efficient features from audio waves. Furthermore, the gate structures of each LSTM unit was optimised in [102] for efficiency.

Autoencoder: An autoencoder comprises an encoder to map the input to hidden representations and a decoder to project the hidden features back to the input data. In [103], a stacked autoencoder was proposed for identifying S1 and S2 sounds, and the trained model outperformed deep belief neural networks as well as classic ML models like SVMs.

B. Deep Learning for Classification

DL employs complex models to learn effective representations directly from raw heart sounds or simple time-frequency representations. In addition to the pipeline shown in Figure 5, this section summarises the advances of DL methods for heart sound classification as follows.

Deep learning on time-frequency representations: Given the challenge of extracting effective representations from raw heart sound signals, 2D time-frequency representations have been widely employed as input for 2D CNNs in heart sound classification [68]. In [104], spectrograms extracted by STFT from heart sounds were fed into ResNet for abnormal heart sounds detection. Multi-domain features were considered more comprehensive in reflecting the characteristics of all heart sound classes. In [105], spectrograms, Mel spectrograms and MFCCs were employed, and the final predictions were obtained through ensemble learning.

In addition to CNNs, which are effective for extracting spatial features, RNNs excel at capturing temporal features. LSTM-RNNs have been applied to process discrete wavelet transforms and MFCCs for heart sound classification [87], [106]. Deng et al. [107] employed convolutional recurrent neural networks (CRNNs), combining CNNs and RNNs, for heart sound classification. CRNNs were also used for detecting murmurs in [108]. In another study, LSTMs were combined with CNNs for abnormal heart sound detection in [109]. Additionally, alternative classifiers have been used for

heart sound classification, including a stacked sparse autoencoder deep neural network [110] and a semi-non-negative matrix factorisation classifier [111].

Deep learning on other features: In addition to the excellent works mentioned above on DL applied to time-frequency representations, other features extracted from heart sounds have also been utilised, including time-domain features, and 1D/2D frequency-based features.

- i) *Time-domain features:* Similar to features used in classic ML approaches, 1D time-domain features can be also fed into DL models for heart sound classification. For instance, the instant energy of heart sounds was extracted as the input for stacked auto-encoder networks in [51]. Additionally, multiple statistical features (e.g., mean, median, and variance) were extracted from all 75 ms segments in each complete heart sound clip and fed into a bidirectional LSTM (BiLSTM)-RNN model for classification in [112].
- ii) *1D frequency-based features:* Either 1D CNNs or feed-forward DNN models can be used to process 1D features. In [113], general frequency features and Mel-domain features were fed into 1D CNNs, and then multiple CNNs were assembled for the final prediction. Mel spectrograms and MFCC were employed to extract additional features, serving as the input for a 5-layer feed-forward DNN model in [8].
- iii) *2D frequency-based features:* Herein, 2D frequency-based features are listed to include (a) multiple 1D frequency features directly extracted from audio segments rather than window functions in the STFT domain and (b) features computed from time-frequency features. Qian et al. utilised wavelets to calculate wavelet energy features from a set of short acoustic segments and further used GRU-RNNs as the classifier [114]. Dong et al. extracted log Mel features and corresponding functionals from heart sound segments and implemented classification LSTM-RNNs and GRU-RNNs [82]. In their experiments, log Mel features performed better than MFCCs and other LLDs [82]. Zhang et al. extracted temporal quasi-periodic features computed by an average magnitude difference function from spectrograms and applied LSTM-RNNs to explore the dependency relation within the features [115]. Additionally, a denoising auto-encoder was employed to extract deep representations from spectrograms as the input of the classifier of 1D CNNs in [116].

End-to-end learning: In recent years, as the selection of time-frequency representations and other features still requires human efforts, there has been a growing trend towards using end-to-end networks to learn representations directly from heart sounds. Various 1D CNN architectures based on raw heart sound signals have been proposed and applied to the task of heart sound classification [9], [117]. Furthermore, Liu et al. introduced a temporal convolutional network (TCN) that exhibited high sensitivity for heart sound classification [118], as a TCN benefiting from dilated and casual convolutions is more suitable for sequential data than typical CNNs.

Additionally, a 1D CNN model consisting of residual blocks was developed for classifying heart sounds [119]. Moreover, GRU-RNNs were used to process raw heart sound signals for the screening of heart failure [120].

Several studies have also highlighted the capability of CNNs and RNNs in learning frequency-domain and time-domain characteristics of heart sounds. For instance, Shuvo et al. proposed a CardioXNet model that employed representation learning followed by sequence residual learning without any preprocessing [121]. In the representation learning phase, three parallel 1D CNN pathways were constructed to extract time-invariant features from heart sound signals. In the sequence residual learning phase, BiLSTM-RNNs were employed to learn sequential representation. The study in [41] attempted to automatically learn time-frequency features. Specifically, frequency-domain features were extracted by 1D CNNs, and the time-domain characteristics were extracted by GRU-RNNs. A self-attention mechanism was further used to fuse the two types of features for the final classification. In [122], time-convolution (tConv) layers were implemented at the front end of the network for learning finite impulse response filters.

Transfer learning: Due to extremely strict regulations governing data collection in the healthcare domain, heart sound datasets are typically not as large as datasets in other areas of Computer Audition. To overcome this limitation, transfer learning has emerged as a valuable approach, employing pre-trained DL models optimised on large-scale datasets. In recent studies, pre-trained models are primarily learnt on either an image dataset (i.e., ImageNet [123]) and an audio dataset (i.e., AudioSet [124]). Although heart sounds are presented as audio signals, a different data type from images, DL models trained on ImageNet have demonstrated good performance on time-frequency representations extracted from heart sounds for heart sound classification. This is attributed to the fact that the large-scale ImageNet dataset improves the generalisation of DL models, and the time-frequency representations can be fed into pre-trained DL models as colourful images. Typical DL models on ImageNet, such as AlexNet [125] and VGG [126], have been successfully repurposed for heart sound classification [7], [35]. Compared to ImageNet, AudioSet includes multiple types of acoustic signals and therefore is more closely related to heart sounds in terms of data type. In [127], pre-trained audio neural networks (PANNs) trained on AudioSet were used for classifying heart sounds with inputs of time-frequency representations. It was found that PANNs outperformed ImageNet-based models [127], including VGG, MobileNet V2, ResNet, and ResNeXt.

After extracting representations using pre-trained models, transfer learning uses various types of classifiers for classification, mainly including classic ML classifiers and feed-forward neural networks. For instance, SVMs were applied to representations extracted by AlexNet, VGG16, and VGG19 in [35], [40]. Additionally, other classifiers such as KNNs were used in [40]. In another approach [7], a

pre-trained VGG model was frozen and followed by fully connected layers.

Additionally, fine-tuning pre-trained models in transfer learning has also shown good performance for heart sound classification. For example, a fine-tuned AlexNet provided effective representations for heart sound classification [35], [40]. Similarly, PANNs were fine-tuned in [127]. Fine-tuned models have even outperformed pre-trained models as they adapt to the data distribution of heart sound datasets. In [7], fine-tuned VGG performed better than pre-trained VGG when SVMs were used as classifiers.

C. Interpretation

Explainable DL has emerged as a crucial topic in healthcare. Developing explainable DL models can foster trust among physicians and patients by providing insights into model predictions. This section categorises interpretation methods into model-agnostic and model-specific approaches. Model-agnostic methods are independent of ML/DL model structures, whereas model-specific methods are closely tied to model architectures.

Model-agnostic interpretation: The shapley additive explanations (SHAP) algorithm [128] is based on Sharpley values, which indicate the feature importance in a prediction model according to the Game Theory. Sharpley values locally explain model predictions for each data sample, while SHAP can provide both local and global interpretations. The study in [129] used SHAP to locally explain a VGG model for heart sound classification. The findings of the study revealed that S1 and S2 heart sounds exhibited high feature importance in six types of time-frequency representations, including STFT, log Mel spectrogram, Hilbert—Huang transformation (HHT), wavelet transform (WT), MFCC, and Stockwell transform (ST). Interestingly, it was observed that low-frequency information in the time-frequency representations positively contributed to predicting normal heart sounds, while high-frequency information had a negative impact. Conversely, this trend was opposite for abnormal heart sounds [129].

S1 and S2 heart sounds were also found important for heart sound classification in [130]. The study [130] compared the SHAP algorithm with the occlusion map visualisation method for model interpretation. The occlusion map evaluates feature importance by masking partial feature regions, offering an alternative perspective on model explanation. It was found that the trained model might still correctly classify the heart sounds when the S1 heart sound was masked by the occlusion map. This observation suggests that other heart sound regions may also contribute to model predictions [130].

Model-specific interpretation: In [131], an attention mechanism was used to visualise the contribution of each feature unit to model predictions. The attention mechanism was applied to CNN, LSTM-RNN, and GRU-RNN models. It was found that the attention heatmap of heart sounds with the moderate/severe state can reveal irregular characteristics, while normal heart sounds with regular heartbeats showed regular feature importance along the time axis in the attention maps.

Similarly, in [108], a temporal attention pooling mechanism was used to assign importance weights to each frame in systolic murmur regions. With the attention mechanism, the murmur regions exhibited high importance for murmurs detection.

V. Published Resources and Advanced Performance

A. Published Datasets

In the past years, several heart sound databases have been collected. The following access-available databases are briefly introduced in Table III.

The PASCAL challenge Database [132] was split into two sets A and B. In dataset A, 176 heart sounds (0.393 hours) were recorded with the iStethoscope Pro iPhone app and annotated into S1 and S2 sounds for heart sound segmentation. Each heart sound in dataset A was also labelled into one of the four classes: *normal*, *murmur*, *extra heart sound*, and *artifact*. Dataset B with 656 recordings (1.194 hours) was annotated into three classes: *normal*, *murmur*, and *extrasystole*.

The PhysioNet/CinC Database [25] used in the PhysioNet/CinC Challenge 2016 [133] consists of multiple databases recorded from different data collectors. The publicly available training set includes five databases collected from both healthy individuals and patients. It comprises 3,240 recordings (20.216 hours in total) from more than 764 subjects. The task was set as a three-class classification task: *normal* and *abnormal*, and *noisy*.

The Heart Sounds Shenzhen (HSS) corpus [82] used in the INTERSPEECH Computational Paralinguistic challenge (ComParE) 2018 was collected by the Shenzhen University General Hospital. The dataset consists of 845 recordings (7.047 hours) from 170 subjects (f: 55, m: 115) using with an electronic stethoscope. Each audio sample was annotated into one of the three classes: *normal*, *mild*, and *moderate/severe*.

A heart sound dataset available on GitHub [134] contains 1,000 audio files (0.679 hours in total). The audio recordings are balanced across five classes: *normal*, *aortic stenosis*, *mitral regurgitation*, *mitral stenosis*, and *mitral valve prolapse*.

The Michigan Heart sound database^{1,2} provides heart sounds from different areas and poses: the apex area when a subject is supine, the apex area for left decubitus, the aortic area when sitting, and the pulmonic area for supine. It consists of 23 heart sound recordings with a total duration of 0.413 hours. The heart sounds were annotated into *normal* and multiple *pathological* states.

The CirCor DigiScope Database [135], used in the George B. Moody PhysioNet Challenge 2022 [136], was collected from a pediatric population aged 21 years or younger. The heart sounds were recorded from one or multiple locations, including pulmonary valve, aortic valve, mitral valve, tricuspid valve, and others. The publicly available training set consists of 3,163 audio

¹[Online]. Available: <https://open.umich.edu/find/open-educational-resources/medical/heart-sound-murmur-library>

²[Online]. Available: https://www.med.umich.edu/lrc/psb_open/html/repo/primer_heartsound/primer_heartsound.html

TABLE III Published datasets for heart sound classification. AS: Aortic stenosis, MR: Mitral regurgitation, MS: Mitral stenosis, MVP: Mitral valve prolapse. Notably, the statistics in this table only considered accessible data sets.

DATASET	CHALLENGE	#SAMPLES	DURATION (h)	#SUBJECTS	TASK
PASCAL Database [132]	PASCAL Challenge [132]	176	0.393	unknown	Dataset A: Normal, Murmur, Extra Heart Sound, Artifact
		656	1.194	unknown	Dataset B: Normal, Murmur, Extrasystole
PhysioNet/CinC Database [25]	PhysioNet/CinC Challenge 2016 [133]	3,240	20.216	764+	Normal, Abnormal, Too noisy or ambiguous
HSS [82]	ComParE Challenge 2018 [69]	845	7.047	170	Normal, Mild, Moderate/Severe
Data on GitHub [134]	–	1,000	0.679	unknown	Normal, AS, MR, MS, MVP
Michigan Heart sound database ³	–	23	0.413	unknown	Normal, Pathological
CirCor DigiScope Database [135]	George B. Moody PhysioNet Challenge 2022 [136]	3,163	20.094	942	Normal, abnormal; presence, absence, or unclear cases of murmurs

samples totaling with 20.094 hours from 942 participants. Two classification tasks were targeted: i) *normal* and *abnormal*, and ii) *presence of murmurs*, *absence of murmurs*, and *unclear cases of murmurs*.

B. State-of-the-Art Performance

As follows, the benchmarks in the challenges associated with the aforementioned databases are discussed. For databases without established benchmarks, the state-of-the-art performance in the challenges is analysed. Additionally, advanced research studies pertaining to databases not used in challenges are reviewed.

The PASCAL database was utilised in the Classifying Heart Sounds Challenge [132]. The champion team in the challenge extracted hand-crafted features based on the segmented S1 and S2 sounds [137], and trained FNNs to classify the heart sounds. On Dataset A, the proposed approach achieved precision values of 0.35, 0.67, 0.18, and 0.92 for normal, murmur, extra-sound, and artifact, respectively. On Dataset B, the method achieved precision values of 0.70, 0.30, and 0.67 for normal, murmur, and extrasystole, respectively.

The PhysioNet/CinC challenge offered a benchmark using selected hand-crafted features and a logistic regression classifier [25]. The features were extracted from segmented four states, and then partial features were selected using logistic regression. The resulting sensitivity and specificity were 0.62 and 0.70, respectively. The highest average score achieved in the challenge was 0.86 (sensitivity: 0.94, specificity: 0.78) [138]. The winning approach utilised features extracted from the above four states, which were then fed into a variant of an AdaBoost classifier. Additionally, heart sounds segmented into cardiac cycles and decomposed into multiple frequency bands were processed by a CNN classifier. Finally, an ensemble of the AdaBoost and the CNN classifiers were used for final predictions.

The HSS database was released with a benchmark in the COMPARE challenge 2018 [69]. An unweighted average score of 0.562 was achieved by fusing the best two models among COMPARE features + SVM, Bag-of-Words features + SVM, and auDeep features + RNN + SVM. The fusion strategy of majority voting outperformed multiple single-model methods.

On the dataset available on GitHub [134], a high accuracy of 0.988 in the five-class classification task was achieved by an SVM classifier using both MFCCs and wavelet transform features. The sensitivity and the specificity were 0.982 and 0.994, respectively. Furthermore, multiple heart valve diseases, including MR, MS, and AS, were distinguished from the healthy control with an accuracy of 0.9833 in [139].

Due to its smaller size compared to other databases, most approaches applied to the Michigan Heart sound database have used hand-crafted features and classic ML classifiers. For instance, the study in [140] achieved an accuracy of 0.8889 using FNNs with hand-crafted features for a nine-class classification. Another study in [86] employed MFCCs and FNNs to correctly classify all samples, identifying 13 types of apex heart sounds.

The CirCor DigiScope Database [135], released in the George B. Moody PhysioNet Challenge 2022 [136], facilitated multiple evaluation metrics, such as F-measure and accuracy. The highest weighted accuracy achieved on the test set [141] was 0.78. This performance was attained by a CNN model using augmented Mel spectrograms as input for classifying heart murmurs.

C. Published Algorithms

Although only a few codes are publicly available in recent years, it is worth noting that the abundance of released codes in 2016 was largely attributed to the PhysioNet challenge 2016 [25]. Notably, codes in 2016 are omitted from this study to focus on the most advanced studies during 2017–2022.

³[Online]. Available: <https://open.umich.edu/find/open-educational-resources/medical/heart-sound-murmur-library> https://www.med.umich.edu/lrc/psb-open/html/repo/primer_heartsound/primer_heartsound.html

In [134], the authors provided a Matlab code⁴ for training deep neural networks utilising multiple features, including MFCCs and features extracted through a discrete wavelet transform. Additionally, the study presented in [122] implemented a Python-based CNN model with time-convolutional units, simulating finite impulse response filters.⁵ Furthermore, ResNets applied to linear and logarithmic spectrogram-image features were implemented in a Python code⁶ shared by the authors of [142]. Lastly, a Matlab code⁷ for detecting valvular heart disease from heart sounds and echocardiograms was released in [143].

VI. Future Research Directions and Open Issues

A. Findings

In classification tasks that primarily focus on the screening for heart diseases, segmentation is often considered as a preprocessing procedure before classification. The question of whether segmentation benefits classification remains open.

Segmentation + Classification: Many studies have employed segmentation techniques or pre-existing segmentation information as a preprocessing step before the classification procedure. For instance, segmented cardiac cycles served as input for DL models in [122]. Clips beginning from the S1 heart sound with a fixed length of 1.6 s were utilised for classification in [120]. The importance of segmentation was demonstrated for abnormal heart sound detection in [130]. Interestingly, experiments in [130] did not show a significant improvement in model performance when segmentation information was incorporated, compared to models without segmentation. This lack of improvement may be attributed to the inherent power and robustness of the models, suggesting that segmentation might be automatically handled by intermediate layers in these models. This assertion was supported by explanations provided by the SHAP algorithm [130]. Additionally, S1 and S2 sounds were observed to be more important compared to other clips within a heart sound segment. Therefore, segmentation is necessary either as an additional procedure for classifiers lacking robustness or as an internal procedure within more advanced classifiers.

No-segmentation: Several approaches have advocated for the use of non-segmented heart sounds, aiming to simplify automated auscultation [9], [80]. Apart from feeding complete heart sound samples into neural networks, heart sounds can be segmented into shorter clips of equal length for model training [105]. For example, the first 5 s of each audio sample were selected as the model input in [6], and segmented 5 s clips were also used in [77]. Most studies employ audio clips with lengths ranging from 2 s to 6 s [41], [45].

B. Limitations and Outlook

Hardware development: In clinics, echocardiography involves obtaining ultrasound scans with a small probe that emits high-frequency sound waves. Physicians can diagnose conditions by observing the heart, blood vessels, and blood flows through this method.⁸ However, echocardiography requires well-trained skills for professionals, limiting its usage in primary care. Classic acoustic stethoscopes used in primary care require physicians and nurses to undergo training. Consequently, there is a high demand for electronic stethoscopes in primary care. In recent years, electronic stethoscopes have been developed to record heart sounds and transmit them to computers or mobile phones for further analysis [26]. Most electronic stethoscopes can only achieve basic functions such as amplifying and visualising heart sounds without providing a diagnosis. More recently, there are a few studies and hardware advancements focused on automated diagnosis. For instance, a field-programmable gate array (FPGA) was designed to classify heart sounds via an LSTM-RNN model in [144]. “HD Steth with ECG”⁹ embedded artificial intelligence (AI) into an electronic stethoscope to detect multiple cardiac abnormalities. As outlined throughout this overview, AI shows promise in diagnosing heart sound abnormalities, thereby reducing the dependence on well-trained professionals. Devices capable of accurately diagnosing cardiac diseases will be invaluable in promoting early screening for cardiac conditions in primary care and home settings.

Performance improvement: Although automated auscultation is ideally expected to replace human analysis, model performance can be a bottleneck for applying automated auscultation to clinical usage. False negative predictions can result in delayed or missed therapies and aggravated conditions. In future efforts, i) automated auscultation will be essential to achieve high performance and should account for individual differences in the context of personalised healthcare. The current research studies are mostly based on heart sounds, while many types of individual information such as demographics can affect model performance [145]. Such individual information can be encoded as inputs for DL models. Additionally, electronic health records (EHR) information can be integrated to prompt personalisation. Heart sound analysis can be implemented based on heart sounds and relevant medical history, thereby providing targeted and timely diagnosis. A long-term dynamic analysis model is also essential for precise diagnosis, taking into account changes in medical status. ii) In terms of ML and DL, the field is currently witnessing the advent and adoption of foundation models [146] (pre-)trained on large-scale datasets. Several approaches using pre-trained models have been observed and listed in this work. However, one can expect even larger models to emerge with the potential for abilities directly related to heart sound analysis as a ‘downstream’ task. On the other hand, the upcoming era of

⁴[Online]. Available: <https://github.com/yaseen21khan/Classification-of-Heart-Sound-Signal-Using-Multiple-Features>

⁵[Online]. Available: <https://github.com/mhealthbuett/heartnet>

⁶[Online]. Available: <https://github.com/mHealthBuet/CepsNET>

⁷[Online]. Available: <https://github.com/uit-hdl/heart-sound-classification>

⁸[Online]. Available: <https://www.nhs.uk/conditions/echocardiogram/>

⁹[Online]. Available: <https://www.stethoscope.com/hd-steth-with-ecg/>

foundation models is expected to be marked by homogenisation, and it remains to be seen if the diversity of heart sound analysis approaches reported herein will indeed converge to a few large data-trained models over the next years [146]. iii) Furthermore, human-machine collaboration holds great promise for improving system performance and providing more accurate diagnoses and timely treatments for patients. Human-machine classification can combine both machine predictions and human input (from crowd workers and experts) to achieve more precise diagnoses [147]. In [147], data samples predicted with high uncertainty were sent to crowd workers for majority voting. Similarly, samples will be forwarded to an expert based on a certainty threshold derived from predictions made by the crowd.

C. Interpretable, Dependable, and Actionable Deep Heart Sound Analysis

The explanation methods in Section IV offer local explanations that interpret DL models on a case-by-case basis, yet lack a global capability of revealing the underlying classification rules or summarising the characteristics of each heart sound class. Explainable DL models, such as deep neural decision trees [148], hold promise for explaining the models themselves from a structure perspective. Learnt or searched data samples of prototypes, criticisms, and counterfactuals [149] can illustrate the typical characteristics of each class, enabling physicians to compare new samples with these heart sounds for improved understanding and analysis. Specifically, by analysing the patterns of criticisms, physicians can potentially reduce the number of false negatives, which is a crucial aspect in the healthcare field. More recently, sonification has been proposed to explain DL models for enhanced human-computer interaction [150]. Unlike visualisation, sonification offers a novel perspective for explaining models through auditory means.

Additionally, considering the health implications, it appears crucial that AI-driven heart sound analysis exhibits the utmost dependability [151]. While mechanisms exist, further adaptation to the specific field of application, if not novel algorithms, will need to be designed. Ultimately, dependability will emerge as a major driving factor for the trustworthiness of heart monitoring solutions. In everyday situations, trustworthiness is a key factor to winning users.

Moreover, to enable DL models to be actionable in real life, data privacy has been another emerging research topic aiming at protecting users' data from leakage or external attacks. Machine unlearning [152] and federated learning methods [153] can help healthcare institutions better organise patients' private data in a secure manner without sacrificing diagnosis accuracy. Furthermore, AI attacks on heart sound analysis, such as through adversarial attacks, needs to be contemplated and dealt with. In summary, DL models hold promise in guiding healthcare providers' actions in their daily practices to provide better care for patients. It will be essential to improve DL models not only in terms of performance but also from human-centred perspectives in future.

D. Real-Life Applications

PCG signals hold promise for applications in abnormal heart sound detection, typically approached as a binary classification task. For predicting specific abnormal heart sounds, several databases, including the PASCAL database [132], HSS [82], and data available on Github [134], have enabled multi-class classification tasks. More specific predictions of heart sounds provide more precise early screening, benefiting both patients and clinicians compared to binary classification. However, the above three databases are not as large as the PhysioNet/CinC database [25] and the CirCor DigiScope database [135] (see Table III). Collecting heart sound data with more detailed labels can be a further step in heart sound analysis, potentially facilitating the release of more databases for relevant research. Furthermore, algorithms such as transfer learning and multi-task learning, which involve both binary classification and multi-class classification, have the potential for achieving more detailed predictions.

In real-life scenarios, automated heart sound analysis holds promise as an early screening tool for patients. The study in [109] presented a pipeline from heart sound recording to abnormal heart sound detection in real-life usage. Once a tool with heart sound analysis is developed, it can be used for real-time heart sound analysis on a daily basis. Herein, recorded data can be processed either online on a cloud server or offline, taking into account users' privacy issues. The study in [153] introduced a federated learning framework for heart sound analysis. Furthermore, users can go to clinics for further diagnosis if abnormal heart sounds are detected. This can help the users promptly seek medical attention, and also reduce the burden on clinics with long waiting lists. In clinics, physicians can benefit from the analysis model's outputs, including predictions and interpretations, to guide the use of suitable diagnosis instruments and provide effective treatments. If the tool is granted permission to monitor patients' heart status, physicians can offer timely and useful suggestions to patients. Additionally, detecting other health statuses from heart sounds shows promise in developing comprehensive healthcare instruments. For example, psychological stress was detected from simultaneous PCG and ECG signals [154].

VII. Conclusion

This work summarised both classic machine learning and deep learning technologies for heart sound analysis from 2017 to 2022, including denoising, segmentation, classification, and interpretation. Available databases were introduced with evaluation metrics in this study. This work also listed publicly available repositories for implementing heart sound classification. Additionally, several findings and limitations of heart sound classification were analysed, and potential future works were discussed. This work presented a summary of the advances in heart sound analysis, provided insightful discussions, and highlighted promising research directions for the community.

Acknowledgment

This work was supported in part by the Federal Ministry of Education and Research (BMBF), Germany through the project LeibnizKILabor under Grant 01DD20003; in part by the Ministry of Science and Technology of the People's Republic of China with the STI2030-Major Projects under Grant 2021ZD0201900; in part by the National Natural Science Foundation of China under Grant 62272044; and in part by the Teli Young Fellow Program from the Beijing Institute of Technology, China. The research was mainly done when Zhao Ren was at the L3S Research Center, Leibniz University Hannover, Germany. The last author further acknowledges help from the Munich Data Science Institute (MDSI) and the Munich Center for Machine Learning (MCML) – both in Munich, Germany.

References

- [1] U. Alam, O. Asghar, S. Q. Khan, S. Hayat, and R. A. Malik, "Cardiac auscultation: An essential clinical skill in decline," *Brit. J. Cardiol.*, vol. 17, pp. 8–10, 2010.
- [2] I. R. Hanna and M. E. Silverman, "A history of cardiac auscultation and some of its contributors," *Amer. J. Cardiol.*, vol. 90, no. 3, pp. 259–267, 2002.
- [3] H. Tang, M. Wang, Y. Hu, B. Guo, and T. Li, "Automated signal quality assessment for heart sound signal by novel features and evaluation in open public datasets," *BioMed Res. Int.*, vol. 2021, pp. 1–15, 2021.
- [4] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.
- [5] B. M. Whitaker, P. B. Suresha, C. Liu, G. D. Clifford, and D. V. Anderson, "Combining sparse coding and time-domain features for heart sound classification," *Physiol. Meas.*, vol. 38, no. 8, pp. 1701–1713, 2017.
- [6] P. Langley and A. Murray, "Heart sound classification from unsegmented phonocardiograms," *Physiol. Meas.*, vol. 38, no. 8, pp. 1658–1670, 2017.
- [7] Z. Ren, N. Cummins, V. Pandit, J. Han, K. Qian, and B. Schuller, "Learning image-based representations for heart sound classification," in *Proc. Int. Conf. Digit. Health*, 2018, pp. 143–147.
- [8] T. H. Chowdhury, K. N. Poudel, and Y. Hu, "Time-frequency analysis, denoising, compression, segmentation, and classification of PCG signals," *IEEE Access*, vol. 8, pp. 160882–160890, 2020.
- [9] B. Xiao, Y. Xu, X. Bi, J. Zhang, and X. Ma, "Heart sounds classification using a novel 1-D convolutional neural network with extremely low parameter consumption," *Neurocomputing*, vol. 392, pp. 153–159, 2020.
- [10] A. K. Bhoi, K. S. Sherpa, and B. Khandelwal, "Multidimensional analytical study of heart sounds: A review," *Int. J. Bioautomation*, vol. 19, no. 3, pp. 351–376, 2015.
- [11] T. Chakrabarti, S. Saha, S. Roy, and I. Chel, "Phonocardiogram signal analysis practices, trends and challenges: A critical review," in *Proc. Int. Conf. Workshop Comput. Commun.*, 2015, pp. 1–4.
- [12] M. Nabih-Ali, E.-S. A. El-Dahshan, and A. S. Yahia, "A review of intelligent systems for heart sound signal analysis," *J. Med. Eng. Technol.*, vol. 41, no. 7, pp. 553–563, 2017.
- [13] S. K. Ghosh, P. R. Nagarajan, and R. K. Tripathy, "Heart sound data acquisition and preprocessing techniques: A review," in *Handbook of Research on Advancements of Artificial Intelligence in Healthcare Engineering*. Hershey PA, USA: IGI Global, 2020, pp. 244–264.
- [14] A. K. Dwivedi, S. A. Imtiaz, and E. Rodriguez-Villegas, "Algorithms for automatic analysis and classification of heart sounds—A systematic review," *IEEE Access*, vol. 7, pp. 8316–8345, 2019.
- [15] G. D. Clifford et al., "Recent advances in heart sound analysis," *Physiol. Meas.*, vol. 38, pp. E10–E25, 2017.
- [16] W. Chen, Q. Sun, X. Chen, G. Xie, H. Wu, and C. Xu, "Deep learning methods for heart sounds classification: A systematic review," *Entropy*, vol. 23, no. 6, 2021, Art. no. 667.
- [17] B. Majhi and A. Kashyap, "Application of soft computing techniques to heart sound classification: A review of the decade," in *Soft Computing Applications and Techniques in Healthcare*. Boca Raton, FL, USA: CRC Press, 2020, pp. 113–138.
- [18] A. Bourrouhou, A. Jilbab, C. Nacir, and A. Hammouch, "Heart sound signals segmentation and multiclass classification," *Int. J. Online Biomed. Eng.*, vol. 16, no. 15, pp. 64–79, 2020.
- [19] V. N. Varghees and K. Ramachandran, "Effective heart sound segmentation and murmur classification using empirical wavelet transform and instantaneous phase for electronic stethoscope," *IEEE Sensors J.*, vol. 17, no. 12, pp. 3861–3872, Jun. 2017.
- [20] H. K. Walker, W. D. Hall, and J. W. Hurst, *Clinical Methods: The History, Physiology, and Laboratory Examinations*. Oxford, U.K.: Butterworths, 1990.
- [21] H. Naseri and M. Homaeinezhad, "Detection and boundary identification of phonocardiogram sounds using an expert frequency-energy based metric," *Ann. Biomed. Eng.*, vol. 41, no. 2, pp. 279–292, 2013.
- [22] C. Ahlström, "Processing of the phonocardiographic signal: Methods for the intelligent stethoscope," Ph.D. dissertation, Dept. Biomedical Eng., Linköping University, Linköping, Sweden, 2006.
- [23] A. M. Noor and M. F. Shadi, "The heart auscultation: From sound to graphical," *J. Eng. Technol.*, vol. 4, no. 2, pp. 73–84, 2013.
- [24] D. S. Gerbarg, A. Taranta, M. Spagnuolo, and J. J. Hofler, "Computer analysis of phonocardiograms," *Prog. Cardiovasc. Dis.*, vol. 5, pp. 393–405, 1963.
- [25] C. Liu et al., "An open access database for the evaluation of heart sound algorithms," *Physiol. Meas.*, vol. 37, no. 12, pp. 2181–2213, 2016.
- [26] S. Leng, R. S. Tan, K. T. C. Chai, C. Wang, D. Ghista, and L. Zhong, "The electronic stethoscope," *Biomed. Eng. Online*, vol. 14, no. 1, pp. 1–37, 2015.
- [27] E. Delgado-Trejos, A. F. Quiceno-Manrique, J. I. Godino-Llorente, M. Blanco-Velasco, and G. Castellanos-Dominguez, "Digital auscultation analysis for heart murmur detection," *Ann. Biomed. Eng.*, vol. 37, no. 2, pp. 337–353, 2009.
- [28] R. J. Martis, U. R. Acharya, and H. Adeli, "Current methods in electrocardiogram characterization," *Comput. Biol. Med.*, vol. 48, pp. 133–149, 2014.
- [29] N. K. Al-Qazzaz, I. F. Abdulazez, and S. A. Ridha, "Simulation recording of an ECG, PCG, and PPG for feature extractions," *Al-Khwarizmi Eng. J.*, vol. 10, no. 4, pp. 81–91, 2014.
- [30] S. M. Shekhatkar, Y. Kotriwar, K. Harikrishnan, and G. Ambika, "Detecting abnormality in heart dynamics from multifractal analysis of ECG signals," *Sci. Rep.*, vol. 7, 2017, Art. no. 15127.
- [31] S. Ari, K. Hembram, and G. Saha, "Detection of cardiac abnormality from PCG signal using LMS based least square SVM classifier," *Expert Syst. Appl.*, vol. 37, no. 12, pp. 8019–8026, 2010.
- [32] S. A. Singh, S. A. Singh, N. D. Devi, and S. Majumder, "Heart abnormality classification using PCG and ECG recordings," *Computación y Sistemas*, vol. 25, no. 2, pp. 381–391, 2021.
- [33] P. S. Molcer, I. Kecskes, V. Deliç, E. Domijan, and M. Domijan, "Examination of formant frequencies for further classification of heart murmurs," in *Proc. IEEE 8th Int. Symp. Intell. Syst. Inform.*, 2010, pp. 575–578.
- [34] Y. Tsao, T. H. Lin, F. Chen, Y. F. Chang, C. H. Cheng, and K. H. Tsai, "Robust S1 and S2 heart sound recognition based on spectral restoration and multi-style training," *Biomed. Signal Process. Control*, vol. 49, pp. 173–180, 2019.
- [35] H. Alaskan, N. Alzhrani, A. Hussain, and F. Almarshed, "The implementation of pretrained AlexNet on PCG classification," in *Proc. Int. Conf. Intell. Comput.*, 2019, pp. 784–794.
- [36] N. M. Khan, M. S. Khan, and G. M. Khan, "Automated heart sound classification from unsegmented phonocardiogram signals using time frequency features," *Int. J. Comput. Inf. Eng.*, vol. 12, no. 8, pp. 598–603, 2018.
- [37] Q. Hu, J. Hu, X. Yu, and Y. Liu, "Automatic heart sound classification using one dimension deep neural network," in *Proc. Int. Conf. Secur., Privacy Anonymity Computation, Commun. Storage*, 2020, pp. 200–208.
- [38] F. Noman, S.-H. Salleh, C.-M. Ting, S. B. Samdin, H. Ombao, and H. Hussain, "A Markov-switching model approach to heart sound segmentation and classification," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 3, pp. 705–716, Mar. 2020.
- [39] Z. Abduh, E. A. Nehary, M. A. Wahed, and Y. M. Kadah, "Classification of heart sounds using fractional fourier transform based mel-frequency spectral coefficients and traditional classifiers," *Biomed. Signal Process. Control*, vol. 57, 2020, Art. no. 101788.
- [40] S. A. Singh, S. Majumder, and M. Mishra, "Classification of short unsegmented heart sound based on deep learning," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf.*, 2019, pp. 1–6.
- [41] S. Li, F. Li, S. Tang, and F. Luo, "Heart sounds classification based on feature fusion using lightweight neural networks," *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, Art. no. 4007009.
- [42] N. Ibrahim, N. Jamal, M. N. A.-H. Sha'abani, and L. F. Mahadi, "A comparative study of heart sound signal classification based on temporal, spectral and geometric features," in *Proc. IEEE-EMBS Conf. Biomed. Eng. Sci.*, 2021, pp. 24–29.
- [43] W. Zhang, J. Han, and S. Deng, "Heart sound classification based on scaled spectrogram and tensor decomposition," *Expert Syst. Appl.*, vol. 84, pp. 220–231, 2017.
- [44] M. Banerjee and S. Majhi, "Multi-class heart sounds classification using 2D-convolutional neural network," in *Proc. 5th Int. Conf. Comput., Commun. Secur.*, 2020, pp. 1–6.
- [45] P. T. Krishnan, P. Balasubramanian, and S. Umapathy, "Automated heart sound classification system from unsegmented phonocardiogram (PCG) using deep neural network," *Phys. Eng. Sci. Med.*, vol. 43, no. 2, pp. 505–515, 2020.
- [46] W. Zeng, J. Yuan, C. Yuan, Q. Wang, F. Liu, and Y. Wang, "A new approach for the detection of abnormal heart sound signals using TQWT, VMD and neural networks," *Artif. Intell. Rev.*, vol. 54, no. 3, pp. 1613–1647, 2021.
- [47] B. Al-Naami, H. Fraihat, N. Y. Gharabeh, and A.-R. M. Al-Hinnawi, "A framework classification of heart sound signals in PhysioNet challenge 2016 using high order statistics and adaptive neuro-fuzzy inference system," *IEEE Access*, vol. 8, pp. 224852–224859, 2020.
- [48] A. I. Humayun et al., "An ensemble of transfer, semi-supervised and supervised learning methods for pathological heart sound classification," in *Proc. Interspeech*, 2018, pp. 127–131.

- [49] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-27, no. 2, pp. 113–120, Apr. 1979.
- [50] M. Baydoun, L. Safatly, H. Ghaziri, and A. E. Haji, "Analysis of heart sound anomalies using ensemble learning," *Biomed. Signal Process. Control*, vol. 62, 2020, Art. no. 102019.
- [51] O. Deperioglu, "Heart sound classification with signal instant energy and stacked autoencoder network," *Biomed. Signal Process. Control*, vol. 64, 2021, Art. no. 102211.
- [52] G. Eslamizadeh and R. Barati, "Heart murmur detection based on wavelet transformation and a synergy between artificial neural network and modified neighbor annealing methods," *Artif. Intell. Med.*, vol. 78, pp. 23–40, 2017.
- [53] M. U. Akram, A. Shaikat, F. Hussain, S. G. Khawaja, and W. H. Butt, "Analysis of PCG signals using quality assessment and homomorphic filters for localization and classification of heart sounds," *Comput. Methods Programs Biomed.*, vol. 164, pp. 143–157, 2018.
- [54] K. A. Babu, B. Ramkumar, and M. S. Manikandan, "Automatic identification of S1 and S2 heart sounds using simultaneous PCG and PPG recordings," *IEEE Sensors J.*, vol. 18, no. 22, pp. 9430–9440, Nov. 2018.
- [55] W. Zhang, J. Han, and S. Deng, "Heart sound classification based on scaled spectrogram and partial least squares regression," *Biomed. Signal Process. Control*, vol. 32, pp. 20–28, 2017.
- [56] M. V. Shervegar and G. V. Bhat, "Heart sound classification using gaussian mixture model," *Porto Biomed. J.*, vol. 3, no. 1, pp. 1–7, 2018.
- [57] M. V. Shervegar and G. V. Bhat, "Automatic segmentation of phonocardiogram using the occurrence of the cardiac events," *Inform. Med. Unlocked*, vol. 9, pp. 6–10, 2017.
- [58] S. E. Schmidt, C. Holst-Hansen, C. Graff, E. Toft, and J. J. Struijk, "Segmentation of heart sound recordings by a duration-dependent hidden Markov model," *Physiol. Meas.*, vol. 31, no. 4, pp. 513–529, 2010.
- [59] S. Das, S. Pal, and M. Mitra, "Acoustic feature based unsupervised approach of heart sound event detection," *Comput. Biol. Med.*, vol. 126, 2020, Art. no. 103990.
- [60] S. Shukla, S. K. Singh, and D. Mitra, "An efficient heart sound segmentation approach using kurtosis and zero frequency filter features," *Biomed. Signal Process. Control*, vol. 57, 2020, Art. no. 101762.
- [61] A. P. Kamson, L. Sharma, and S. Dandapat, "Multi-centroid diastolic duration distribution based HMM for heart sound segmentation," *Biomed. Signal Process. Control*, vol. 48, pp. 265–272, 2019.
- [62] J. Oliveira, F. Renna, and M. Coimbra, "A subject-driven unsupervised hidden semi-Markov model and Gaussian mixture model for heart sound segmentation," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 2, pp. 323–331, May 2019.
- [63] J. Oliveira, F. Renna, T. Mantadelis, and M. Coimbra, "Adaptive sojourn time HMM for heart sound segmentation," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 2, pp. 642–649, Mar. 2019.
- [64] J. Rubin et al., "Recognizing abnormal heart sounds using deep learning," in *Proc. IJCAI Knowl. Discov. Healthcare Workshop*, 2017, pp. 13–19.
- [65] D. M. Nogueira, C. A. Ferreira, E. F. Gomes, and A. M. Jorge, "Classifying heart sounds using images of motifs, MFCC and temporal features," *J. Med. Syst.*, vol. 43, no. 6, 2019, Art. no. 168.
- [66] Y. Chen, S. Wei, and Y. Zhang, "Classification of heart sounds based on the combination of the modified frequency wavelet transform and convolutional neural network," *Med. Biol. Eng. Comput.*, vol. 58, no. 9, pp. 2039–2047, 2020.
- [67] D. B. Springer, L. Tarasenko, and G. D. Clifford, "Logistic regression-HMM-based heart sound segmentation," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 4, pp. 822–832, Apr. 2016.
- [68] A. Duggento, A. Conti, M. Guerri, and N. Toschi, "Classification of real-world pathological phonocardiograms through multi-instance learning," in *Proc. 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2021, pp. 771–774.
- [69] B. Schuller et al., "The INTERSPEECH 2018 computational paralinguistics challenge: Atypical & self-assessed affect, crying & heart beats," in *Proc. Interspeech*, 2018, pp. 122–126.
- [70] F. Eyben et al., "The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing," *IEEE Trans. Affect. Comput.*, vol. 7, no. 2, pp. 190–202, Apr.–Jun. 2016.
- [71] T. Tuncer, S. Dogan, R.-S. Tan, and U. R. Acharya, "Application of Petersen graph pattern technique for automated detection of heart valve diseases with PCG signals," *Inf. Sci.*, vol. 565, pp. 91–104, 2021.
- [72] S. Amiriparian, M. Schmitt, N. Cummins, K. Qian, F. Dong, and B. Schuller, "Deep unsupervised representation learning for abnormal heart sound classification," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2018, pp. 4776–4779.
- [73] S. A. Singh and S. Majumder, "Classification of unsegmented heart sound recording using KNN classifier," *J. Mechanics Med. Biol.*, vol. 19, no. 04, 2019, Art. no. 1950025.
- [74] E. Soares, P. Angelov, and X. Gu, "Autonomous learning multiple-model zero-order classifier for heart sound classification," *Appl. Soft Comput.*, vol. 94, 2020, Art. no. 106449.
- [75] J. Dastagir, F. A. Khan, M. S. Khan, and K. N. Khan, "Computer-aided phonocardiogram classification using multidomain time and frequency features," in *Proc. Int. Conf. Artif. Intell.*, 2021, pp. 50–55.
- [76] S. Khaled, M. Fakhry, H. Esmail, A. Ezzat, and E. Hamad, "Analysis of training optimisation algorithms in the NARX neural network for classification of heart sound signals," *Int. J. Sci. Eng. Res.*, vol. 13, no. 2, pp. 382–390, 2022.
- [77] V. Arora, R. Leekha, R. Singh, and I. Chana, "Heart sound classification using machine learning and phonocardiogram," *Modern Phys. Lett. B*, vol. 33, no. 26, 2019, Art. no. 1950321.
- [78] A. Yadav, M. K. Dutta, C. M. Travieso, and J. B. Alonso, "Automatic classification of normal and abnormal PCG recording heart sound recording using Fourier transform," in *Proc. IEEE Int. Work Conf. Bioinspired Intell.*, 2018, pp. 1–9.
- [79] M. Sotaquirá, D. Alvear, and M. Mondragón, "Phonocardiogram classification using deep neural networks and weighted probability comparisons," *J. Med. Eng. Technol.*, vol. 42, no. 7, pp. 510–517, 2018.
- [80] A. M. Alqudah, "Towards classifying non-segmented heart sound records using instantaneous frequency based features," *J. Med. Eng. Technol.*, vol. 43, no. 7, pp. 418–430, 2019.
- [81] Y. Tan et al., "Heart sound classification based on fractional Fourier transformation entropy," in *Proc. IEEE 4th Glob. Conf. Life Sci. Technol.*, 2022, pp. 588–589.
- [82] F. Dong et al., "Machine listening for heart status monitoring: Introducing and benchmarking HSS—The heart sounds Shenzhen corpus," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 7, pp. 2082–2092, Jul. 2020.
- [83] V. Maknickas and A. Maknickas, "Recognition of normal-abnormal phonocardiographic signals using deep convolutional neural networks and mel-frequency spectral coefficients," *Physiol. Meas.*, vol. 38, no. 8, pp. 1671–1684, 2017.
- [84] J. Chen, X. Dang, and M. Li, "Heart sound classification method based on ensemble learning," in *Proc. Int. Conf. Intell. Comput. Signal Process.*, 2022, pp. 8–13.
- [85] J. F. Chen and X. Dang, "Heart sound analysis based on extended features and related factors," in *Proc. IEEE Symp. Ser. Comput. Intell.*, 2019, pp. 2189–2194.
- [86] M. Rahmandani, H. A. Nugroho, and N. A. Setiawan, "Cardiac sound classification using mel-frequency cepstral coefficients (MFCC) and artificial neural network (ANN)," in *Proc. 3rd Int. Conf. Inf. Technol., Inf. System Elect. Eng.*, 2018, pp. 22–26.
- [87] F. A. Khan, A. Abid, and M. S. Khan, "Automatic heart sound classification from segmented/unsegmented phonocardiogram signals using time and frequency features," *Physiol. Meas.*, vol. 41, no. 5, 2020, Art. no. 055006.
- [88] M. Adiban, B. BabaAli, and S. Shehnepoor, "Statistical feature embedding for heart sound classification," *J. Elect. Eng.*, vol. 70, no. 4, pp. 259–272, 2019.
- [89] S. R. Thiyagaraja et al., "A novel heart-mobile interface for detection and classification of heart sounds," *Biomed. Signal Process. Control*, vol. 45, pp. 313–324, 2018.
- [90] J. Li, L. Ke, Q. Du, X. Ding, X. Chen, and D. Wang, "Heart sound signal classification algorithm: A combination of wavelet scattering transform and twin support vector machine," *IEEE Access*, vol. 7, pp. 179339–179348, 2019.
- [91] N. Mei, H. Wang, Y. Zhang, F. Liu, X. Jiang, and S. Wei, "Classification of heart sounds based on quality assessment and wavelet scattering transform," *Comput. Biol. Med.*, vol. 137, 2021, Art. no. 104814.
- [92] S. K. Ghosh, R. K. Tripathy, R. Ponnalagu, and R. B. Pachori, "Automated detection of heart valve disorders from the PCG signal using time-frequency magnitude and phase features," *IEEE Sens. Lett.*, vol. 3, no. 12, Dec. 2019, Art. no. 7002604.
- [93] N. K. Sawant, S. Patidar, N. Nesaragi, and U. R. Acharya, "Automated detection of abnormal heart sound signals using fano-factor constrained tunable quality wavelet transform," *Biocybernetics Biomed. Eng.*, vol. 41, no. 1, pp. 111–126, 2021.
- [94] M. E. Karar, S. H. El-Khafif, and M. A. El-Brawany, "Automated diagnosis of heart sounds using rule-based classification tree," *J. Med. Syst.*, vol. 41, no. 4, pp. 1–7, 2017.
- [95] R. F. Ibarra-Hernández, N. Bertin, M. A. Alonso-Arévalo, and H. A. Guillén-Ramírez, "A benchmark of heart sound classification systems based on sparse decompositions," *Proc. SPIE*, vol. 10975, pp. 26–38, 2018.
- [96] F. Renna, J. Oliveira, and M. T. Coimbra, "Deep convolutional neural networks for heart sound segmentation," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 6, pp. 2435–2445, Nov. 2019.
- [97] E. Messner, M. Zöhrer, and F. Pernkopf, "Heart sound segmentation—An event detection approach using deep recurrent neural networks," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 1964–1974, Sep. 2018.
- [98] Y. Chen, J. Lv, Y. Sun, and B. Jia, "Heart sound segmentation via duration long-short term memory neural network," *Appl. Soft Comput.*, vol. 95, 2020, Art. no. 106540.
- [99] T. Fan, J. Zhu, Y. Cheng, Q. Li, D. Xue, and R. Munnoch, "A new Direct heart sound segmentation approach using bi-directional GRU," in *Proc. 24th Int. Conf. Automat. Comput.*, 2018, pp. 1–5.
- [100] T. Fernando, H. Ghaemmaghami, S. Denman, S. Sridharan, N. Hussain, and C. Fookes, "Heart sound segmentation using bidirectional LSTMs with attention," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 6, pp. 1601–1609, Jun. 2020.
- [101] Y. Chen, Y. Sun, J. Lv, B. Jia, and X. Huang, "End-to-end heart sound segmentation using deep convolutional recurrent network," *Complex Intell. Syst.*, vol. 7, pp. 2103–2117, 2021.
- [102] C. Xu, J. Zhou, L. Li, J. Wang, D. Ying, and Q. Li, "Heart sound segmentation based on SMGU-RNN," in *Proc. Third Int. Conf. Biol. Inf. Biomed. Eng.*, 2019, pp. 1–7.

- [103] M. Mishra, H. Menon, and A. Mukherjee, "Characterization of S.1 and S.2 heart sounds using stacked autoencoder and convolutional neural network," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 9, pp. 3211–3220, Sep. 2019.
- [104] A. Balamurugan, S. G. Teo, J. Yang, Z. Peng, Y. Xulei, and Z. Zeng, "ResHNNet: Spectrograms based efficient heart sounds classification using stacked residual networks," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Inform.*, 2019, pp. 1–4.
- [105] J. M.-T. Wu et al., "Applying an ensemble convolutional neural network with Savitzky–Golay filter to construct a phonocardiogram prediction model," *Appl. Soft Comput.*, vol. 78, pp. 29–40, 2019.
- [106] B. Ahmad, F. A. Khan, K. N. Khan, and M. S. Khan, "Automatic classification of heart sounds using long short-term memory," in *Proc. Int. Conf. Open Source Syst. Technol.*, 2021, pp. 1–6.
- [107] M. Deng, T. Meng, J. Cao, S. Wang, J. Zhang, and H. Fan, "Heart sound classification based on improved MFCC features and convolutional recurrent neural networks," *Neural Netw.*, vol. 130, pp. 22–32, 2020.
- [108] J.-K. Wang et al., "Automatic recognition of murmurs of ventricular septal defect using convolutional recurrent neural networks with temporal attentive pooling," *Sci. Rep.*, vol. 10, 2020, Art. no. 21797.
- [109] S. Pandya, T. R. Gadekallu, P. K. Reddy, W. Wang, and M. Alazab, "InfusedHeart: A novel knowledge-infused learning framework for diagnosis of cardiovascular events," *IEEE Trans. Comput. Social Syst.*, early access, Mar. 2, 2022, doi: 10.1109/TCSS.2022.3151643.
- [110] Z. Abduh, E. A. Nehary, M. A. Wahed, and Y. M. Kadah, "Classification of heart sounds using fractional Fourier transform based mel-frequency spectral coefficients and stacked autoencoder deep neural network," *J. Med. Imag. Health Inform.*, vol. 9, no. 1, pp. 1–8, 2019.
- [111] W. Han, S. Xie, Z. Yang, S. Zhou, and H. Huang, "Heart sound classification using the SNMFNet classifier," *Physiol. Meas.*, vol. 40, no. 10, 2019, Art. no. 105003.
- [112] M. Fakhry and A. F. Brery, "A comparison study on training optimization algorithms in the biLSTM neural network for classification of PCG signals," in *Proc. Int. Conf. Innov. Res. Appl. Sci., Eng. Technol.*, 2022, pp. 1–6.
- [113] K. Ranipa, W.-P. Zhu, and M. Swamy, "Multimodal CNN fusion architecture with multi-features for heart sound classification," in *Proc. IEEE Int. Symp. Circuits Syst.*, 2021, pp. 1–5.
- [114] K. Qian, Z. Ren, F. Dong, W.-H. Lai, B. Schuller, and Y. Yamamoto, "Deep wavelets for heart sound classification," in *Proc. Int. Symp. Intell. Signal Process. Commun. Syst.*, 2019, pp. 1–2.
- [115] W. Zhang, J. Han, and S. Deng, "Abnormal heart sound detection using temporal quasi-periodic features and long short-term memory without segmentation," *Biomed. Signal Process. Control*, vol. 53, 2019, Art. no. 101560.
- [116] F. Li et al., "Feature extraction and classification of heart sound using 1D convolutional neural networks," *EURASIP J. Adv. Signal Process.*, vol. 2019, no. 1, pp. 1–11, 2019.
- [117] R. Avanzato and F. Beritelli, "Heart sound multiclass analysis based on raw data and convolutional neural network," *IEEE Sens. Lett.*, vol. 4, no. 12, Dec. 2020, Art. no. 7004104.
- [118] K. Liu, L. Yuan, C. Huang, W. Wu, Q. Wang, and G. Wu, "Abnormal heart sound detection by using temporal convolutional network," in *Proc. IEEE Asia-Pacific Conf. Image Process., Electron. Comput.*, 2022, pp. 1026–1029.
- [119] S. L. Oh et al., "Classification of heart sound signals using a novel deep WaveNet model," *Comput. Methods Programs Biomed.*, vol. 196, 2020, Art. no. 105604.
- [120] S. Gao, Y. Zheng, and X. Guo, "Gated recurrent unit-based heart sound analysis for heart failure screening," *Biomed. Eng. Online*, vol. 19, no. 1, pp. 1–17, 2020.
- [121] S. B. Shuvo, S. N. Ali, S. I. Swapnil, M. S. Al-Rakhami, and A. Gumaie, "CardioXNet: A novel lightweight deep learning framework for cardiovascular disease classification using heart sound recordings," *IEEE Access*, vol. 9, pp. 36955–36967, 2021.
- [122] A. I. Humayun, S. Ghaffarzadegan, M. I. Ansari, Z. Feng, and T. Hasan, "Towards domain invariant heart sound abnormality detection using learnable filterbanks," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 8, pp. 2189–2198, Aug. 2020.
- [123] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 248–255.
- [124] J. F. Gemmeke et al., "Audio set: An ontology and human-labeled dataset for audio events," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2017, pp. 776–780.
- [125] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017.
- [126] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Representations*, 2015, pp. 1–14.
- [127] T. Koike, K. Qian, Q. Kong, M. D. Plumbley, B. W. Schuller, and Y. Yamamoto, "Audio for audio is better? An investigation on transfer learning models for heart sound classification," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2020, pp. 74–77.
- [128] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. Annu. Conf. Neural Inf. Process. Syst.*, 2017, pp. 4768–4777.
- [129] Z. Wang, K. Qian, H. Liu, B. Hu, B. W. Schuller, and Y. Yamamoto, "Exploring interpretable representations for heart sound abnormality detection," *Biomed. Signal Process. Control*, vol. 82, 2023, Art. no. 104569.
- [130] T. Dissanayake, T. Fernando, S. Denman, S. Sridharan, H. Ghaemmaghami, and C. Fookes, "A robust interpretable deep learning classifier for heart anomaly detection without segmentation," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 6, pp. 2162–2171, Jun. 2021.
- [131] Z. Ren et al., "Deep attention-based neural networks for explainable heart sound classification," *Mach. Learn. Appl.*, vol. 9, 2022, Art. no. 100322.
- [132] P. Bentley, G. Nordehn, M. Coimbra, and S. Mannor, "The PASCAL classifying heart sounds challenge 2011 (CHSC2011) Results." [Online]. Available: <http://www.peterjbentley.com/heartchallenge/index.html>
- [133] A. Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
- [134] G.-Y. Son and S. Kwon, "Classification of heart sound signal using multiple features," *Appl. Sci.*, vol. 8, no. 12, 2018, Art. no. 2344.
- [135] J. Oliveira et al., "The CirCor DigiScope dataset: From murmur detection to murmur classification," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 6, pp. 2524–2535, Jun. 2022.
- [136] M. A. Reyna et al., "Heart murmur detection from phonocardiogram recordings: The George B. Moody PhysioNet challenge 2022," *PLOS Digit. Health*, vol. 2, no. 9, pp. 1–22, 2023.
- [137] E. F. Gomes and E. Pereira, "Classifying heart sounds using peak location for segmentation and feature construction," in *Proc. Workshop Classifying Heart Sounds*, 2012, pp. 480–92.
- [138] C. Potes, S. Parvaneh, A. Rahman, and B. Conroy, "Ensemble of feature-based and deep learning-based classifiers for detection of abnormal heart sounds," in *Proc. Comput. Cardiol. Conf.*, 2016, pp. 621–624.
- [139] S. K. Ghosh, R. Ponnalagu, R. Tripathy, and U. R. Acharya, "Heart sound feature extraction and classification using autoregressive power spectral density (AR-PSD) and statistics features," in *Proc. Int. Conf. Sci. Technol.*, vol. 1755, 2015, Art. no. 090007.
- [140] D. Kristomo, R. Hidayat, I. Soesanti, and A. Kusjani, "Heart sound feature extraction and classification using autoregressive power spectral density (AR-PSD) and statistics features," *AIP Conf. Proc.*, vol. 1755, 2016, Art. no. 090007.
- [141] H. Lu et al., "A lightweight robust approach for automatic heart murmurs and clinical outcomes classification from phonocardiogram recordings," in *Proc. Comput. Cardiol. Conf.*, 2022, pp. 1–4.
- [142] F. B. Azam et al., "Cardiac anomaly detection considering an additive noise and convolutional distortion model of heart sound recordings," *Artif. Intell. Med.*, vol. 133, pp. 1–12, 2022.
- [143] P. N. Waaler et al., "Algorithm for predicting valvular heart disease from heart sounds in an unselected cohort," *Frontiers Cardiovascular Med.*, pp. 1–15, 2024.
- [144] W.-S. Jhong et al., "Deep learning hardware/software co-design for heart sound classification," in *Proc. Int. SoC Des. Conf.*, 2020, pp. 27–28.
- [145] W. R. Thompson, A. J. Reinisch, M. J. Unterberger, and A. J. Schrieffer, "Artificial intelligence-assisted auscultation of heart murmurs: Validation by virtual clinical trial," *Pediatr. Cardiol.*, vol. 40, no. 3, pp. 623–629, 2019.
- [146] R. Bommasani et al., "On the opportunities and risks of foundation models," 2021, *arXiv:2108.07258*.
- [147] W. Callaghan, J. Goh, M. Mohareb, A. Lim, and E. Law, "Mechanicalheart: A human-machine framework for the classification of phonocardiograms," *Proc. ACM Hum.-Comput. Interaction*, vol. 2, no. CSCW, pp. 1–17, 2018.
- [148] Y. Yang, I. G. Morillo, and T. M. Hospedales, "Deep neural decision trees," in *Proc. Int. Conf. Mach. Learn. WHI*, 2018, pp. 34–40.
- [149] A. V. Looveren and J. Klaise, "Interpretable counterfactual explanations guided by prototypes," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discov. Databases*, 2021, pp. 650–665.
- [150] B. W. Schuller et al., "Towards sonification in multimodal and user-friendly explainable artificial intelligence," in *Proc. Int. Conf. Multimodal Interaction*, 2021, pp. 788–792.
- [151] M. Pelillo and T. Scantamburlo, *Machines We Trust: Perspectives on Dependable AI*. Cambridge, MA, USA: MIT Press, 2021.
- [152] V. Gupta, C. Jung, S. Neel, A. Roth, S. Sharifi-Malvajerd, and C. Waites, "Adaptive machine unlearning," in *Proc. Annu. Conf. Neural Inf. Process. Syst.*, 2021, pp. 788–792.
- [153] W. Qiu et al., "A federated learning paradigm for heart sound classification," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2022, pp. 1045–1048.
- [154] A. Cheema and M. Singh, "Psychological stress detection using phonocardiography signal: An empirical mode decomposition approach," *Biomed. Signal Process. Control*, vol. 49, pp. 493–505, 2019.