Noisy Neonatal Chest Sound Separation for High-Quality Heart and Lung Sounds

Ethan Grooby ©, Student Member, IEEE, Chiranjibi Sitaula ©, Davood Fattahi ©, Reza Sameni ©, Senior Member, IEEE, Kenneth Tan, Lindsay Zhou, Arrabella King, Ashwin Ramanathan, Atul Malhotra ©, Guy Dumont ©, Life Fellow, IEEE, and Faezeh Marzbanrad ©, Senior Member, IEEE

Abstract—Stethoscope-recorded chest sounds provide the opportunity for remote cardio-respiratory health monitoring of neonates. However, reliable monitoring requires high-quality heart and lung sounds. This paper presents novel artificial intelligence-based Non-negative Matrix Factorisation (NMF) and Non-negative Matrix Co-Factorisation (NMCF) methods for neonatal chest sound separation. To assess these methods and compare them with existing single-channel separation methods, an artificial mixture dataset was generated comprising heart, lung, and noise sounds. Signal-to-noise ratios were then calculated for these artificial mixtures. These methods were also tested on real-world noisy neonatal chest sounds and assessed based on vital sign estimation error, and a signal quality score of 1-5, developed in our previous works. Overall, both the proposed NMF and NMCF methods outperform the next best existing method by 2.7 dB to 11.6 dB for the artificial dataset, and 0.40 to 1.12 signal quality improvement for the real-world dataset. The median processing time for the sound separation of a 10 s recording was found to be 28.3 s for NMCF and 342 ms for NMF. With the stable and robust performance of our proposed methods, we believe these methods are useful to denoise neonatal heart and lung sounds in the real-world environment.

Index Terms—Artificial intelligence, breath sound, heart sound, lung sound, neonatal, phonocardiogram (PCG), signal quality, single-channel sound separation, telehealth.

I. INTRODUCTION

A CCURATE and timely assessment for serious health problems such as cardio-respiratory diseases is an essential requirement to provide care to newborns [1]. Recording chest sounds with a stethoscope is a common method to obtain such information. In recent times, the availability of digital stethoscopes for neonates has attracted several studies, mostly to enhance the cardio-respiratory sound information obtained, but with the ability to connect with smartphones for mobile health, it may also enable remote healthcare through telehealth. These studies have looked into how to obtain accurate heart and breathing rates [2], predict respiratory distress at birth [3], detect changes in sound characteristics after surfactant treatment, and analyse the sound differences in different neonatal populations [4], [5], [6]. However, a higher level of noise in the neonatal intensive care environment in comparison to adult wards, has resulted in poor quality chest sound recordings and inaccurate assessment [2], [7], [8].

Noise in stethoscope-recorded chest sounds can be broken up into four groups. Firstly, there are external noises, such as crying, stethoscope movement, talking, and general background noise in neonatal intensive care [9]. Secondly, the diagnostically important heart and lung sounds act as noise sources for one another. Thirdly, internal body sounds such as bowel sounds, gastric reflux, and air swallowing. Finally, neonates, particularly ones born earlier than 32 weeks, can experience conditions requiring respiratory support such as high flow, ventilator or bubble Continuous Positive Airway Pressure (CPAP), which are major sources of interference. Overall, it is essential to reduce these noises, and separate heart and lung sounds before any assessment and diagnosis.

Our study focuses on single-channel sound separation and denoising to obtain high-quality heart and lung sounds. Single-channel refers to only a single chest sound recording with no other signals being recorded. Table I summarises existing single-channel sound separation methods and Section III compares...
them. Many of the methods utilise heart sound segmentation to obtain S1 and S2 peaks, which are the first and second heart sounds respectively, that occur due to heart valve closures during the cardiac cycle. As shown in our past works, heart sound segmentation accuracy drops significantly for low-quality recordings [2], [7]. Overall this means most methods are only effective in low-noise scenarios. For lung sound separation, this has proven difficult due to its large frequency band which overlaps with heart and noise sources.

Another limitation of past works is that they rely on adult-based parameters, which are not suitable for the separation of neonatal chest sounds. To address this, existing methods that rely on heart sound segmentation utilise a modified version suitable for neonates, which was developed in our past work [2]. Additionally, instead of single-value parameters being utilised, a set of relevant parameters suitable for neonatal chest sound separation are considered and tested, as highlighted in Tables I and II. Finally, for SSA (i) specifically, an additional constraint of top eigenvalue pairs having the strongest frequency component less than 250 Hz being assigned as heart components is added. This constraint was added to avoid misclassification of lung components as heart.

There has also been research on a deep learning-based autoencoder for sound separation based on adult chest sounds [24]. However, successfully implementing it in the neonatal population requires a large and diverse set of neonatal chest sound recordings, which is currently not available.

This paper presents new approaches based on Non-negative Matrix Factorisation (NMF) and Non-negative Matrix Co-Factorisation (NMCF) to obtain high-quality heart and lung sounds. Different from past works, the model is trained on a fixed small training set of high-quality heart, lung, and noise sound examples. In NMF this training occurs beforehand, whereas in NMCF this training occurs in parallel with separating the noisy recording into heart, lung, and noise sounds. In both cases, this reference set is independent of the validation and test data. The training set enables the utilisation of more detailed information about the frequency and temporal aspects of heart, lung, and noise. Whereas for NMCF, the parallel training and separation of sounds from the noisy recordings enables the model to adapt more specifically to that recording. Overall, this method provides higher quality heart and lung sounds to be generated for further analysis.

A preliminary version of our NMCF work has been reported, which utilised high-quality heart and lung sounds, with no reference noise examples [16]. Initial results on a real-world dataset showed it was superior to existing NMF methods [16]. Four key contributions are presented in this paper. First, existing single-channel denoising, and heart and lung sound separation methods originally developed for adults, were adapted and implemented on newborn chest sound recordings. Second, the new NMF and NMCF approaches specifically for the newborn population are proposed. Third, we incorporate a noise

### TABLE I

| Method (letter reference) | Description |
|--------------------------|-------------|
| Adaptive Fourier Decomposition (c) | Based on the energy distribution, adaptive Fourier decomposition is used to clean identified S1 and S2 heart sound peaks [2], [10], [11], Top 5, 10, 15, 20, and 25 decomposition levels assigned as heart sounds are tested. |
| Adaptive Line Enhancement (d) | Applies adaptive filtering on the original and time-delayed recording to extract semi-periodic components (heart sounds) [12]. Time delay values of 1, 10, 40, 100, 200, and 400 samples are tested. |
| EMD (e) | Decomposes signal into a sum of oscillatory functions called intrinsic mode functions (IMFs). For each IMF, S1 and S2 peaks are identified and filtered to obtain heart and lung components [13]. Empirical mode decomposition (EMD), ensemble EMD, and complete ensemble EMD with adaptive noise are tested. |
| Filtering (f) | Passband frequencies of 50-250 Hz and 200-1000 Hz were used to obtain heart and lung sounds, respectively [2], [7]. |
| Interpolation (g) | Identified S1 and S2 heart sound peaks are removed either by 20-300Hz bandstop filter or complete elimination of them sections [2], [11], [14]. Interpolation in the time-frequency domain is then performed to recover the lung sounds [14]. |
| Modulation Filtering (h) | Involves bandpass and bandstop filtering of temporal trajectories of the short-term spectral components, to obtain heart and lung sounds respectively [15]. The filter ranges tested are 1-20Hz, 2-20Hz, 3-20Hz, 4-20Hz, 5-20Hz, and 6-20Hz. |
| NMF Clustering 1 and 2 (i and j) | Both methods blindly decompose the mixture into numerous sub-components. All components are then clustered into either heart or lung based on spectral or temporal criteria [16]–[18]. |
| Recursive Least Squares Adaptive Filtering (k) | Identified S1 and S2 heart sound peaks are used to create a reference heart sound [2], [11], [19]. The reference least squares filter then uses the original recording and the reference heart sound to obtain clean heart and lung sounds [19]. |
| SSA (l) | Singular spectrum analysis (SSA) decomposes the signal into principal components [20]. Top eigenvector pairs of components with the strongest frequency component less than 250 Hz are assigned as heart sounds [20], [21]. |

### TABLE II

| Method | Best Parameter(s) |
|--------|------------------|
| Adaptive Fourier Decomposition (c) | Decomposition level of 3 (Cry Noise), and 10 (No Noise, Stethoscope Movement Noise, Respiratory Support Noise) |
| Adaptive Line Enhancement (d) | Delay of 1 (No Noise, Respiratory Support Noise), 10 (Stethoscope Movement Noise), and 30 (Cry Noise) |
| EMD (e) | Decompose using ensemble EMD (All Cases) |
| Interpolation (g) | Remove entire segments containing heart sounds and interpolate (All Cases) |
| Modulation Filtering (h) | Bandstop of 3-20Hz (All Cases) for heart sounds and Bandpass of 4-20Hz (No Noise, Ventilator CPAP and General Noise) and 6-20Hz (Bubble CPAP Noise) for lung sounds |
| Wavelet Transform-based Filter (m) | Adaptive threshold of 3 (No Noise, Respiratory Support Noise, Cry Noise), and 3.5 (Stethoscope Movement Noise) |

*Letter reference is used in Section III, Figure 3, and Table IV as shorthand for method names.*
component in the NMF and NMCF models trained on common noise sources for newborns, to separate the sounds into not only heart and lung sounds, but also noise sounds. Finally, the methods are assessed using artificial and real-world noisy neonatal chest sounds.

II. METHODS

A. Non-Negative Matrix Factorisation and Non-Negative Matrix Co-Factorisation

NMF decomposes a given non-negative matrix \( V \in \mathbb{R}^{F \times T} \), into two non-negative matrices \( W \in \mathbb{R}^{K \times F} \) and \( H \in \mathbb{R}^{K \times T} \) (1), where \( K < \min(F; T) \) and \( E \in \mathbb{R}^{F \times T} \) represents the reconstruction error between \( V \) and \( WH \).

\[
V = WH + E \quad (1)
\]

In denoising and sound separation, \( V \) represents the magnitude of the time-frequency transform of the recording mixture, where time is represented as columns and frequency as rows [16]. Examples of time-frequency transforms typically used in NMF are Short-Time Fourier Transform (STFT), Q-transform, and Gammatone filterbank [16], [25].

The basis matrix \( W \), contains the basis column vectors \( w_1 \) to \( w_K \) that represent the spectral pattern of different types of signal sources (e.g., heart, lung, and noise) or their sub-components. The activation matrix \( H \), contains the temporal activation row vectors \( h_1 \) to \( h_K \), which represents when the signal sources occur during a particular time frame.

For fully- and semi-supervised NMF, the basis matrix \( W \) is optimised with pure and synchronously recorded reference heart, lung, and noise sounds during the training phase. \( W \) is then fixed and \( H \) is optimised with the mixture recording during the test phase. In NMCF, instead of having a separate training and test phase, \( W \) is optimised simultaneously with sound separation. This method enables more efficient sound separation, as the mixture recording can also contribute to the training of \( W \) [16].

The reference heart, lung, and noise sounds are represented by \( b_h, b_l, \) and \( b_n \), number of components respectively. The basis column vectors \( w_1 \) to \( w_K \) and activation row vectors \( h_1 \) to \( h_K \), can be combined such that that the first set of components (1 to \( b_h \)), second set of components (\( b_h + 1 \) to \( b_h + b_l \)) and third set of components (\( b_h + b_l + 1 \) to \( b_h + b_l + b_n \)) represent heart \((V_h = W_hH_h)\), lung \((V_l = W_lH_l)\), and noise \((V_n = W_nH_n)\) respectively [16]. Therefore, \( V \) can be explicitly separated into \( V = V_h + V_l + V_n \).

As obtaining pure and synchronously recorded sounds is not feasible, we propose a new modified version of NMF (3) and NMCF (4). In this version, datasets of high-quality heart, lung, and noise sounds are used in the cost function to enable the generalisation of \( W_h \), \( W_l \) and \( W_n \) respectively for the sound separation. Note, that these reference datasets are not obtained from the same subject as the noisy mixture recording that is being denoised. Additionally, the aim is that reference datasets generalise well, such that they can be used on a variety of noisy chest sound recordings, and still produce high-quality heart and lung sounds. This aim was assessed in Section II-D and the results are shown in Section III.

In both the proposed NMF (3) and NMCF (4), the matrices \( W \) and \( H \) are optimised by minimising the cost function \( D (2). D_\beta \) refers to \( \beta \)-divergence cost function, with the most popular values being \( \beta = 0, 1, \) or 2 [16]. A sparsity penalty on the activation matrix \( H \) is calculated based on the L1-norm of \( H \), and \( \mu \) controls the importance of the sparsity constraint. The sparsity penalty enables more detailed decomposition both temporally and spectrally, while ensuring only a small set of meaningful basis vectors are active at a single time frame [26].

\[
D(V|\hat{W}H) = D_\beta(V|\hat{W}H) + \mu||H||_1, \quad \text{for} \quad \beta \in [0, 1]
\]

For the proposed NMF method (3), we have introduced a cost function that utilises datasets of clean heart, lung, and noise sounds that are obtained from different subjects than the noisy mixture recording. This differs from past work which either requires simultaneous reference recordings from the same subject or relies on blind decomposition [16], [27].

\[
W_h = \min_{W_h, H_h} \left( \sum_{ih=1}^{eh} D(V_h^{(ih)}|\hat{W}_hH_h^{(ih)}) \right)
\]

\[
W_l = \min_{W_l, H_l} \left( \sum_{il=1}^{el} D(V_l^{(il)}|\hat{W}_lH_l^{(il)}) \right)
\]

\[
W_n = \min_{W_n, H_n} \left( \sum_{in=1}^{en} D(V_n^{(in)}|\hat{W}_nH_n^{(in)}) \right)
\]

\[
H_m = \min_{W_m, H_m} (D(V_m|\hat{W}H_m))
\]

Where: \( H_m = [H_{mh}; H_{ml}; H_{mn}; H_{mun}], \)

\[
\hat{W} = [\hat{W}_h, \hat{W}_l, \hat{W}_n, \hat{W}_{um}]
\]

Note: \( ; \) = new column and, \( = \) new row within a matrix

\( eh, el, \) and \( en \) are the total number of heart, lung, and noise examples used in the reference database, whereas \( ih, il, \) and \( in \) index the specific example being used. \( V_h^{(ih)} \) is the magnitude of time-frequency representation and \( H_h^{(ih)} \) is the activation matrix for reference heart example \( ih \). Similarly for \( V_l^{(il)} \) and \( H_l^{(il)} \) for lung example \( il \), as well as, \( V_n^{(in)} \) and \( H_n^{(in)} \) for noise example \( in \). \( H_m \) is the activation matrix of the mixture recording that needs to be denoised, which is broken into four components, heart \((H_{mh})\), lung \((H_{ml})\), noise \((H_{mn})\) and unsupervised component \((H_{mun})\). Similarly for basis matrix \( W \) for \( \hat{W}_h, \hat{W}_l, \hat{W}_n, \) and \( \hat{W}_{um} \). The unsupervised component \( \hat{W}_{um} \) is added.
to deal with the large variety of noises that are not covered in the reference noise dataset, avoiding these components being assigned to the heart or lung components.

For the proposed NMCF method (4), building on our past work, we have introduced a supervised noise component [16]. The weighting factors \( \lambda_h, \lambda_l, \) and \( \lambda_n \) represent the level of co-factorisation, and are treated as hyperparameters.

\[
W, H_m = \min_{W, H_m, H_h, H_l, H_n} \left( \lambda_h D(V_m|\hat{W}_h H_{mh}) + \lambda_l D(V_m|\hat{W}_l H_{ml}) + \lambda_n D(V_m|\hat{W}_n H_{mn}) 
+ D(V_m|\hat{W}_{un} H_{mun}) 
+ \frac{1}{eh} \sum_{ih=1}^{eh} D(V_h^{(ih)}|\hat{W}_h H_{h}^{(ih)}) 
+ \frac{1}{el} \sum_{il=1}^{el} D(V_l^{(il)}|\hat{W}_l H_{l}^{(il)}) 
+ \frac{1}{en} \sum_{in=1}^{en} D(V_n^{(in)}|\hat{W}_n H_{n}^{(in)}) \right)
\]

Where:
\[
H_m = [H_{mh}; H_{ml}; H_{mn}; H_{mun}],
\hat{W} = [\hat{W}_h, \hat{W}_l, \hat{W}_n, \hat{W}_{un}]
\]

(4)

Based on the cost function in (4), the multiplicative update rule for \( W \) and \( H \) are shown in (5) and (6) respectively. The multiplicative update rule ensures the non-negative constraint on both \( W \) and \( H \). Note that division and \( \otimes \) refer to element-wise division and multiplication. Further details on the derivation of multiplicative update rule from cost function can be found in [16] and [26]. The update rules are then applied in Algorithm: Proposed NMCF Method, to separate the mixture recording into heart and lung sounds, with the overall workflow shown in Fig. 1.

\[
W \leftarrow W \otimes W_{num}(V, W, H) / W_{dem}(V, W, H)
\]

Where:
\[
A = WH, \quad I = F by F matrix of ones,
\]

\[
W_{num}(V, W, H) = (\Lambda^{\beta-2} \otimes V)H^T
+ \hat{W} \otimes (1(\hat{W} \otimes (\Lambda^{\beta-1}H^T))),
\]

\[
W_{dem}(V, W, H) = \Lambda^{\beta-1}H^T
+ \hat{W} \otimes (1(\hat{W} \otimes ((\Lambda^{\beta-2} \otimes V)H^T))
\]

(5)

\[
H \leftarrow H \otimes H_{num}(V, W, H) / H_{dem}(V, W, H)
\]

Where:
\[
A = WHH,
\]

\[
H_{num}(V, W, H) = \hat{W}^T(V \otimes A^{-2}),
\]

\[
H_{dem}(V, W, H) = \hat{W}^T \Lambda^{-1} + \mu
\]

(6)

B. Data Acquisition and Preprocessing

The study was conducted at Monash Newborn, Monash Children’s Hospital, and approved by the Monash Health Human Research Ethics Committee (HREA/18/MonH/471). One-minute recordings were obtained from the right anterior chest of preterm and term newborns using a digital stethoscope (CliniCloud Stethoscope) [5], [6], [7], [28]. The recordings were saved by a commercially available smartphone software in MP3 format with a 16 kHz sampling rate, and then downsampled to 4 kHz for future analysis [29].

A subset of these recordings had synchronous vital signs, i.e., heart and breathing rate for Section II-D2. These vital signs were obtained directly from the bedside monitor, based on the electrocardiogram for heart rate and electrical impedance tomography for breathing rate. In total, 397 recordings were obtained without synchronous vital signs, which included 79 recordings with the newborn on respiratory support. Additionally, 31 recordings were obtained with synchronous vital signs, which included 9 recordings with the newborn on respiratory support. These recordings were then utilised in Section II-D as shown in Fig. 2.

C. Reference Sounds

Reference heart, lung, crying, stethoscope movement, and respiratory support sounds were required for two purposes. Firstly, several existing methods and the proposed NMF and NMCF methods required reference high-quality sounds to enable sound separation. Secondly, these reference sounds were used to create artificial mixtures to enable evaluation of the sound separation methods as described in Section II-D1.

Heart and lung sounds were obtained from the recordings of newborns without respiratory support. As pure heart and lung sounds are required to construct an artificial dataset, these recordings were 4th-order Butterworth bandpass filtered with passband frequencies 50–250 Hz and 200–1000 Hz, to separate heart and lung sounds, respectively [2], [27].

The filtered recordings were annotated by 3 clinicians and 4 electrical engineers familiar with biomedical auscultation for heart and lung signal quality on a 5-level scale. A score of 1 referred to noisy and hardly detectable heartbeats/breathing periods, and 5 referred to clear heart/lung sounds with little to no noise. Mean annotated scores of 4 and above were assessed visually and through audio, and only the recordings with strong heart/lung sounds and little to no noise were chosen as reference signals. In total, 17 signals (9 subjects) with 7 (3 subjects) having synchronous heart rate remained for reference heart sounds, and 9 signals (7 subjects) with no synchronous breathing rate remained for reference lung sounds.

To obtain the cry sounds, regions containing at least 10 s of crying were determined and extracted using our previously developed cry detection algorithm [2]. These segments were then 2nd-order Butterworth high-pass filtered with cutoff 300 Hz to remove heart sounds. Finally, regions not containing crying such as inhale and other lung sounds were replaced with zeros. In total 42 signals (41 subjects) were obtained.

Two clinicians and 1 electrical engineer familiar with biomedical auscultation manually annotated recordings for the presence...
GROOBY et al.: NOISY NEONATAL CHEST SOUND SEPARATION FOR HIGH-QUALITY HEART AND LUNG SOUNDS 2639

Fig. 1. Proposed NMCF Method. (1) Recordings are converted into the time-frequency domain using the short-time Fourier transform (STFT). (2) Variables/Matrices are initialised with random values between zero and one. (3) NMCF is performed with reference clean noise, heart, and lung sounds to obtain the basis matrix $W$ and activation matrix $H$, with corresponding components of heart, lung, noise, and unsupervised. (4/5) Individual components are separated and inverse short-time Fourier transform (ISTFT) with phase is used to obtain heart, lung, and noise (supervised and unsupervised components) sounds.

Part 1: Convert Audio to Time-Frequency Domain
1. $V_{n0} \text{ Phase} = \text{STFT} (\text{audio}_{n0})$ (Mixture Recording)
2. $V^{(h)}_{n0} = \text{STFT} (\text{audio}_{n0}^{(h)})$ (Reference Heart Sounds)
3. $V^{(l)}_{n0} = \text{STFT} (\text{audio}_{n0}^{(l)})$ (Reference Lung Sounds)
4. $V^{(n)}_{n0} = \text{STFT} (\text{audio}_{n0}^{(n)})$ (Reference Noise Sounds)

Part 2: Initialise Variables
5. initialisation: $H_{n0}^{(h)}, H_{n0}^{(l)}, H_{n0}^{(n)}, H_{n0}^{(un)}$
6. $W = [W_{h}, W_{l}, W_{n}, W_{un}]$ (Basis Matrix)
7. $H_{n0}^{(i)} = [H_{n0}^{(h)}, H_{n0}^{(l)}, H_{n0}^{(n)}, H_{n0}^{(un)}]$ (Activation Matrix of Mixture Recording)

Part 3: Training and Test Phase using (5) and (6)
8. for $i=1$ to $\text{max iterations}$ do
9. $H_{h}^{(i)} = H_{h}^{(i)} \oplus \text{normalisation} (W_{h})$
10. $H_{l}^{(i)} = H_{l}^{(i)} \oplus \text{normalisation} (W_{l})$
11. $H_{n}^{(i)} = H_{n}^{(i)} \oplus \text{normalisation} (W_{n})$
12. $H_{un}^{(i)} = H_{un}^{(i)} \oplus \text{normalisation} (W_{un})$
13. $W_{h} = W_{h} \oplus \lambda_{w} (V_{n0}^{(h)} - W_{h} H_{n0}^{(h)})$
14. $W_{l} = W_{l} \oplus \lambda_{w} (V_{n0}^{(l)} - W_{l} H_{n0}^{(l)})$
15. $W_{n} = W_{n} \oplus \lambda_{w} (V_{n0}^{(n)} - W_{n} H_{n0}^{(n)})$
16. $W_{un} = W_{un} \oplus \lambda_{w} (V_{n0}^{(un)} - W_{un} H_{n0}^{(un)})$
17. end for

Part 4: Obtain Separated Heart and Lung Sounds
22. $\text{mask}_{h} = W_{h} H_{n0}^{(h)}$
23. $\text{mask}_{l} = W_{l} H_{n0}^{(l)}$
24. $V_{h} = V_{n0} \oplus \text{mask}_{h}$
25. $V_{l} = V_{n0} \oplus \text{mask}_{l}$

Part 5: Convert Back to Audio (Time Domain)
26. $\text{audio}_{heart} = \text{ISTFT} (V_{h}, \text{Phase})$
27. $\text{audio}_{lung} = \text{ISTFT} (V_{l}, \text{Phase})$

Fig. 2. Usage of Data Flowchart. Within the boxes is the number of recordings, and within the brackets is the number of subjects. *Stmov = Stethoscope movement.

and volume (low, medium, or high) of stethoscope movement and respiratory support noise. Stethoscope movement noise can be typically characterised as a short $< 2$ s burst of noise. These sections of stethoscope movement were isolated and classified as either disconnection or rubbing noise. Disconnection noise only has the stethoscope movement sound present and not heart nor lung sounds, as the stethoscope has disconnected from the newborn’s chest. Whereas for rubbing noise, the stethoscope is still in connection with the newborn’s chest, and can potentially still contain heart and lung sounds. In total, 18 signals (12
subjects) of varying lengths were obtained. To ensure all signals were of length 10 s, zeros were appended to the beginning and end of the signals. The ratio of the total number of zeros placed at the beginning or end of the signals was randomly determined, to ensure that stethoscope movement occurs at a random position in the 10 s signals.

Recordings annotated with a high level of respiratory support noise and minimal other sounds were chosen. From this, 2 signals (1 subject) of bubble CPAP and 5 signals (1 subject) of ventilator CPAP were obtained. Pure respiratory support sounds were also collected by placing the digital stethoscope on the respiratory support machine or tubing. An additional 3 signals for both bubble and ventilator CPAP were obtained from the respiratory support machine itself.

D. Performance Evaluation

1) Artificial Mixtures Evaluation: For the assessment of sound separation methods, generating artificial heart-lung sound mixtures is common [10], [12], [17], [18], [20], [24]. We aim to not only create artificial heart-lung sound mixtures, but also to incorporate relevant real-world noise sounds. Using the reference sounds obtained in Section II-C, artificial mixtures are generated. The most common method in literature for generating chest sound mixtures is the simple addition of all reference sounds, referred to as an instantaneous mixture (7). However, as chest sounds are not simple instantaneous mixtures of heart, lung, and noise sounds, a convolutive mixture model was also adopted (8) [20]. To achieve convolutive mixing, three randomly generated finite-impulse response (FIR) filters of length 4 were adopted (8) [20]. To achieve convolutive mixing, three randomly generated finite-impulse response (FIR) filters of length 4 were adopted (8) [20]. To achieve convolutive mixing, three randomly generated finite-impulse response (FIR) filters of length 4 were adopted (8) [20]. To achieve convolutive mixing, three randomly generated finite-impulse response (FIR) filters of length 4 were adopted (8) [20]. To achieve convolutive mixing, three randomly generated finite-impulse response (FIR) filters of length 4 were adopted (8) [20]. To achieve convolutive mixing, three randomly generated finite-impulse response (FIR) filters of length 4 were adopted (8) [20].

\[
s_{\text{mixture}}(t) = s_{\text{heart}}(t) + s_{\text{lung}}(t) + s_{\text{noise}}(t) \quad (7)
\]

\[
s_{\text{mixture}}(t) = \sum_{k=0}^{3} (a_{\text{heart}}(k)s_{\text{heart}}(t-k) + a_{\text{lung}}(k)s_{\text{lung}}(t-k) + a_{\text{noise}}(k)s_{\text{noise}}(t-k)) \quad (8)
\]

Before mixing heart, lung, and noise sounds, they were scaled to achieve the desired signal-to-noise ratio. Heart sounds were first scaled to achieve a heart-to-lung sounds ratio of $-10, -5, 0, 5, 10, 15,$ and $20$ dB (9). Once scaled, heart and lung sounds were combined and then scaled to achieve chest-to-noise sounds ratio of $-10, -5, 0, 0, 10, 15,$ and $20$ dB (10).

\[
s_{\text{chest}}(t) = 10^{\frac{\text{factor}}{20}} \cdot s_{\text{heart}}(t) + s_{\text{lung}}(t) \quad (9)
\]

\[
s_{\text{mixture}}(t) = 10^{\frac{\text{factor}}{20}} \cdot s_{\text{chest}}(t) + s_{\text{noise}}(t) \quad (10)
\]

Using the generated artificial mixtures and reference sounds used to create these mixtures, signal quality metrics can be calculated using the blind source separation evaluation toolbox [30], [31]. With this toolbox, estimated heart, lung, and noise sounds are decomposed into four components, namely; true reference sound ($s_{\text{target}}$), interference noise ($e_{\text{inter}}$), additive noise ($e_{\text{noise}}$), and algorithmic artifact noise ($e_{\text{artif}}$)

Once decomposed, signal-to-distortion ratio (SDR (11)), signal-to-interference ratio (SIR (12)) and scale-invariant signal-to-distortion ratio (SI-SDR (11)) are calculated. Both SDR and SI-SDR are overall metrics of signal quality. SDR uses a full 512-tap FIR filter, whereas SI-SDR uses a single coefficient to account for allowable scaling discrepancies between estimated separated sounds and reference sounds [30], [32]. Therefore SI-SDR harshly penalises temporal distortions, and is only suitable for the evaluation of instantaneous mixtures in comparison to SDR.

\[
\text{SDR} = 10\log_{10} \frac{||s_{\text{target}}(t)||^2}{||e_{\text{inter}}(t) + e_{\text{noise}}(t) + e_{\text{artif}}(t)||^2} \quad (11)
\]

\[
\text{SIR} = 10\log_{10} \frac{||s_{\text{target}}(t)||^2}{||e_{\text{inter}}(t)||^2} \quad (12)
\]

To construct the artificial mixture dataset, the reference sounds were separated into seven mutually exclusive groups, ensuring recordings from the same subject were within the same group, that is, subject-wise seven-fold cross-validation. One fold was used for hyperparameter optimisation for the proposed NMCF and NMF methods, as well as existing sound separation methods, as shown in Tables I, II, and III, and Section II-E. The remaining six-folds were used for the evaluation of the sound separation methods, with results shown in Fig. 3(a) and 3(b).

2) Heart Rate and Breathing Rate Error: A goal of obtaining high-quality heart and lung sounds is to achieve accurate heart and breathing rate estimates. These vital sign estimates are essential in cardio-respiratory health assessment, to enable proper clinical care to be determined and provided [8], [33].

For the heart audio recordings, heart rate in beats per minute was estimated for each second with a sliding window of 3 s. Heart rate was calculated using the modified version of the method by Springer et al. [11] for a heart rate range of 70-220 beats per minute, as proposed in our past work [7].

For lung audio recordings, breathing rate in breaths per minute was estimated every second with a sliding window of 6 s.

| Parameter | Time-Frequency Representation | Parameter Values |
|-----------|-----------------------------|------------------|
| STFT, FFT Size = 1024 samples, Window Size = 512 samples and Hop Size = 256 samples | SDR | 1.0 (All Cases) |
| NMCF, All Cases, NMF No Noise, General Noise, Bubble CPAP Noise. | SIR | 0.1 (All Cases) |
| STFT, FFT Size = 2048 samples, Window Size = 2048 samples and Hop Size = 512 samples | SI-SDR | 0.25 (No Noise) |
| NMCF, Ventilator CPAP Noise | SI-SDR | 0.25 (Stethoscope Movement Noise, Ventilator CPAP Noise) |
| NMCF, Bubble CPAP Noise | SI-SDR | 0.75 (Cry Noise, Bubble CPAP Noise) |
Artificial Dataset Sound Separation Results. The top 9 subplots are for heart sound separation (Fig. 3(a)), and the bottom 9 subplots are for lung sound separation (Fig. 3(b)). SDR, SIR, and SI-SDR improvement results are calculated according to Section II-D1, and displayed using box and violin plots as described in Fig. 4. No noise refers to only heart-lung sound mixtures, general noise refers to heart-lung sound and either cry or stethoscope movement noise mixtures, and respiratory support noise refers to heart-lung sound and either bubble or ventilator CPAP noise mixtures. Methods a and b (in blue) are the proposed NMCF and NMF methods, and methods c to n (in black) are existing methods, as specified in Table I.
The breathing rate was calculated from the 300–450 Hz power spectral envelope, as this is the region of the strongest lung sound signal. Peak detection is then performed with a breathing rate range of 15 to 100 breaths per minute [2].

3) Signal Quality Assessment: An automated signal quality assessment method to classify real-world heart and lung signal quality on a 5-level scale was developed in our previous works [2], [7]. A score of 1 referred to only noisy and hardly detectable heartbeats/breathing periods, and 5 referred to clear heart/lung sounds with little to no noise.

To calculate the signal quality, 400 features were extracted from chest sound recordings. These features included statistical features (variance, skewness, and kurtosis), predictive fitting coefficients, heart and lung segmentation quality and agreement, Mel-frequency coefficients, wavelet, entropy, and power [7]. Then using the maximum relevance minimum redundancy feature selection algorithm, up to 15 features for heart signal quality, and up to 20 features for lung signal quality, were used as input into a support vector machine (SVM) regression classifier. The SVM classifier had a radial basis function kernel and automatic kernel scale.

In total, the regression classifier was trained on 206 recordings from 97 subjects for lung sound quality estimation, and 223 recordings from 92 subjects for heart sound quality estimation. Reference quality annotations were provided by 3 clinicians and 4 electrical engineers familiar with biomedical auscultation [2], [7]. Note that these recordings are from different subjects than the reference sounds used for training the sound separation methods and creation of the artificial dataset in Section II-C.

4) Real-Time Analysis: The median time for chest sound separation for an example 10 s segment was calculated using MATLAB 2021a with MacBook Pro CPU 2.3 GHz 8-Core Intel i9.

Computation cost per 10 s recording is shown in Table IV. The proposed NMCF method takes a median of 28.2 s and 28.3 s with and without supervised decomposition components respectively. These computational times make the proposed NMCF method not suitable for real-time processing. For the proposed NMF method, the median computational time of 275 ms and 342 ms with and without supervised decomposition components are observed. As the computational times are less than 400 ms, the proposed NMF method is suitable for real-time processing using the stated laptop specifications [16]. For existing methods, adaptive Fourier decomposition (c), adaptive line enhancement (d), filtering (f), interpolation (g), and modulation filtering (h) are suitable for real-time processing using the stated laptop specifications.

5) Statistical Analysis: Statistical tests were performed to determine if the proposed NMF and NMCF methods are significantly outperforming existing methods. Using the Jarque-Bera test, artificial dataset signal quality improvement results were not normally distributed. Therefore, median values are reported and a one-sided Wilcoxon signed-ranked test was used to test significance in Section III. Similarly, vital sign estimation error results were not normally distributed and thus one-sided Wilcoxon signed-rank test was used to test significance. However, as median results were predominately zero, mean and standard deviation results were shown to be more informative. Whereas, signal quality improvement values for the real-world dataset were normally distributed. Therefore, mean and standard deviation values are reported in Table IV and a one-sided t-test was used to test significance in Section III.

E. Implementation

For both the proposed NMCF and NMF, the number of bases used is 20 for the heart ($b_h$), lung ($b_l$), and all noise ($b_n$) sounds.
TABLE V
SUMMARY ARTIFICIAL DATASET SOUND SEPARATION RESULTS. NMCF (a) AND NMF (b) METHODS ARE COMPARED WITH THE BEST PERFORMING EXISTING METHOD WITH MEDIAN IMPROVEMENT RESULTS SHOWN

| Method                      | Heart SDR | Lung SDR | SDR 2 |
|-----------------------------|-----------|----------|-------|
| No Noise                    | a 1.5 b 0.7 | a 0.9 b 0.8* | a 3.1 b 1.4 |
| General Noise- Cry          | a 4.3 b 4.0 | a 8.3 b 6.7 | a 7.7 b 6.0 |
| General Noise- Stethoscope Movement | a 1.7 b 2.1 | a 0.3 b 0.7 | a 0.5 b 0.9 |
| Respiratory Support Noise- Bubble CPAP | a 1.8 b 1.4 | a 0.6 b 0.0** | a 1.9 b 0.4 |
| Respiratory Support Noise- Ventilator CPAP | a 1.8 b 1.0 | a 1 b 0.1 | a 1.2 b 0.9 |

Green shading means improvement and red shading means disimprovement. All results are significant (p-VALUE<0.05) except for * (p-VALUE=0.06) and ** (p-VALUE=0.29).

SDR 2 refers to hard-to-separate lung sound cases, where lung SNR<10 dB for no noise case, and noisy SNR<0 dB for general and respiratory support noise cases.

For heart sound separation, in the no noise case, the proposed NMCF (a) and NMF (b) methods outperformed all existing methods except adaptive line enhancement (d), which had a median SDR improvement of 1.5 dB and 0.8 dB over the proposed methods respectively. However, adaptive line enhancement (d) produced minor temporal distortions in the separated heart sound, resulting in significantly lower SI-SDR values of 18.1 dB and 17.9 dB compared to proposed NMCF (a) and NMF (b) methods respectively. For both general noise (cry and stethoscope movement) and respiratory support noise (bubble CPAP and ventilator CPAP) cases, the proposed methods significantly outperformed all the existing methods.

As both the proposed NMCF (a) and NMF (b) methods do not produce temporal distortions, SI-SDR results are comparable with SDR. Both methods significantly outperformed existing methods in all situations, with median SI-SDR improvement ranging from 5.2 dB for respiratory support noise to 9.9 dB for general noise, compared to the best existing method.

For SIR results, both proposed methods significantly outperformed existing methods in all situations, with median SIR improvement ranging from 3.8 dB for respiratory support noise to 8.2 dB for general noise compared to the next best existing method. For heart sound separation, EMD (e) significantly outperformed both proposed methods in the no noise situation and was comparable for the general noise case. For lung sound separation, the filtering (f) method was comparable to the proposed NMF method for no noise case, but was significantly inferior for hard-to-separate situations (lung SNR less than -10 dB). For all other scenarios, both proposed methods significantly outperformed existing methods for heart and lung sound separation.

Table IV shows the real-world chest sound separation results, as detailed in Sections II-D2 and II-D3. Overall for heart and lung sound separation, all proposed methods significantly outperformed existing methods in all situations. Mean signal quality improvement ranged from 0.40-0.63 for no respiratory support case, and 0.74-1.12 for respiratory support case compared to the next best existing method.

For heart sound separation, all proposed methods outperformed existing methods. Mean signal quality improvement each. For real-world sounds, an unsupervised noise component is also added with the number of bases equaling 10. At most 10 examples of heart (eh), lung (el), and noise (en) sounds, and 100 iterations of multiplicative update (max_iterations) of activation and basis matrix were used.

The following parameters for the proposed NMCF and NMF were tested:

- Time-Frequency Representation:
  - Gammatone filterbank [25] with a frequency range of 50-2000 Hz and frequency bin size of either 128, 256, 512 or 1028.
  - Q-transform with 64 frequency bins per octave
  - STFT with Fast Fourier Transform (FFT) Size = 2048 samples, Window Size = 512 samples and Hop Size = 256 samples similar to previous work [16], or, FFT Size = 1024 samples, Window Size = 512 samples and Hop Size = 256 samples.
- Sparsity penalty (μ): 0, 0.01 and 0.1
- Beta loss (β): 0, 1 and 2
- NMCF λβ, λ1 and λn: 0, 0.25, 0.5, 0.75 and 1

III. RESULTS

Tables II and III show the best parameters for existing and proposed methods based on artificial dataset SDR values, as described in Section II-D1.

Fig. 3 and Table IV show the artificial and real-world chest sound separation results respectively. Methods (a), (a un), (b), and (b un) are the proposed NMCF, NMCF with unsupervised components, NMF, and NMF with unsupervised components methods respectively. Methods (c) to (n) are existing methods, as specified in Table I.

Figs. 3(a) and 3(b), and Table V show the artificial mixture sound separation results for heart and lung sounds, as detailed in Section II-D1. Overall, for heart and lung sound separation, both the proposed NMCF (a) and NMF (b) methods significantly outperformed existing methods in all situations, with median SDR improvement ranging from 2.7 dB for respiratory support noise to 11.6 dB for general noise compared to next best existing method.
compared to the next best existing method ranged between 0.10-0.17 for no respiratory support case, and 0.22-0.38 for the respiratory support case. All signal quality improvement values were significant except the proposed NMCF (a) with the no respiratory support case (p-value = 0.06).

For lung sound separation, all proposed methods significantly outperformed existing methods except filtering (f). Mean signal quality improvement compared to the filtering method (f) ranged from -0.04-0.14 for the no respiratory support case and 0.00-0.27 for the respiratory support case. For no respiratory case, only the proposed NMF method (b) was significantly better than the filtering method (f). Whereas for the respiratory case, both the proposed NMCF (a) and NMCF with unsupervised components (a un) significantly outperformed the filtering method (f).

As seen in Table IV vital sign improvement results were very sporadic. In general, all methods had a median vital sign error improvement of 0, but were positively skewed, as seen in the mean values. Due to the large variation in results, many of the comparisons between methods were insignificant. For no respiratory support, both NMCF methods, (a) and (a un), performed best for heart rate estimation (not significant for NMF clustering 2 (j) and filtering (f)). Whereas NMF without supervised components (b un) performed best for breathing rate estimation (not significant for filtering (f) and modulation filtering (h)). For the presence of respiratory support noise, the existing filtering method (f) performed significantly better than all proposed methods for heart rate, and wavelet transform-based filter (m) performed significantly better than proposed NMCF methods, (a) and (a un).

IV. Discussion

Overall, both the proposed NMCF and NMF methods performed well, producing comparable or superior results compared to existing methods. In the artificial dataset, the proposed NMCF method had a median SDR improvement of 0.00, 1.71, -0.55, 1.17 and 0.92 dB for no noise, cry noise, stethoscope movement noise, bubble CPAP and ventilator CPAP noise, respectively in comparison to the proposed NMF method. Whereas in the real-world dataset, the proposed NMF method outperformed the proposed NMCF method, in the no respiratory support case; conversely, this was the opposite in the respiratory support case. In the presence of hard-to-separate noise such as respiratory support noise, co-factorisation in the proposed NMCF method aids in the better adaption for effective sound separation. NMCF allows for better adaptation of the basis vectors, as opposed to fixed basis vectors in the NMF method, enabling a better ability to segregate noise. As the proposed NMF method can be implemented in real-time, it would be recommended to utilise this method initially. If the signal quality is still poor due to hard-to-separate noise, then the proposed NMCF method should be utilised afterwards.

Both the proposed NMF and NMCF methods produce superior results to existing methods with a relatively small dataset of reference sounds (at most 10 examples per sound type). These example sounds have already been obtained (see Section II-C), and have been used on both artificial mixture sound separation and unseen real-world noisy neonatal chest sound separation with promising results in these varied scenarios. For the artificial dataset, by splitting the reference sounds using subject-wise seven-fold cross-validation, we ensured that the training and test sets are different and diverse.

For the real-world noisy neonatal chest sound separation, an unsupervised component was added to deal with the noise sources that are not represented in the reference noise sounds. The introduction of the unsupervised noise component led to an improvement in signal quality results for NMCF, and a decrease in signal quality for the NMF method (see Table IV). For the proposed NMCF method, the benefit of the unsupervised noise component is that it prevents co-factorisation and learning of the basis vectors and temporal activation to be heavily affected by noise that was not present in the reference sound set. Instead, other noise sources are allocated to the unsupervised component, as they are not related to any of the reference sounds. For the proposed NMF method, the basis matrix has already been pre-computed and fixed for the sound separation phase. Hence, this method is not as affected by the presence of other sources of noise, especially if they have fairly distinct frequency components. Instead, relevant heart and lung sound components can be misplaced into the unsupervised components due to the unsupervised nature of calculating these basis vectors, resulting in a decrease in performance. This is a key limitation with having unsupervised components, as whilst it can remove unknown noise sources, it can also remove some of the heart and lung sound components.

For the reference noise dataset, the most common types of noise sources that are problematic (overlapping frequency band) for obtaining heart and lung sounds were used. However, there are two main concerns with the current dataset. Firstly, unknown noise sources that have frequency overlap with heart and lung sounds. Secondly, the existing noise sounds for cry, stethoscope movement, and respiratory support change in different environments and recording situations. Further work in obtaining chest sounds in a variety of situations and assessing the sound separation method is therefore required. It should be highlighted, that our work is one of the first to model real-world noise sources (not just heart-lung sound mixtures) and quantitatively assess existing and proposed methods to obtain high-quality heart and lung sounds.

Whilst our proposed sound separation methods worked better or comparably to existing methods for respiratory support noise, there is still a large room for improvement. In both the artificial dataset and real-world recordings, respiratory support noise was still present to some extent in the separated sounds. Due to the large frequency overlap with the desired heart and lung sounds, the frequency basis vectors are quite similar for the respiratory support noise, and heart and lung sounds. Some possible ways to improve sound separation in the presence of respiratory support noise, is to have additional reference sounds and add a temporal constraint or decomposition. Currently, only 5 and 8 reference sounds were provided for bubble and ventilator CPAP respectively, which may not provide the entire variety of different flow rates and sounds that can occur.
Vital sign error results were less conclusive and promising for the proposed methods. For median vital sign error improvement being 0, in general, the vital sign estimation methods are designed to be robust to an extent to noise [11]. Instead, the sound separation and denoising methods mainly only improved results when there was a major vital sign estimation error in the raw recordings, explaining the large standard deviation and positive mean seen in the results. For sound separation in the presence of respiratory support noise, the proposed NMF and NMCF methods were inferior to existing methods. One possible explanation for this is that after separation, with respiratory support noise partially removed, there were cases where heart and lung sounds were not present, resulting in an underestimation of vital signs. Whereas for other methods that do not explicitly remove the respiratory support sounds, these noise sounds were discounted as heartbeats/breathing periods, avoiding the underestimation of vital signs. These conclusions are further supported by the existing methods that poorly estimated vital signs in no respiratory support case showing improvement in the more difficult respiratory support case. In particular for lung sounds, all methods struggled to successfully obtain clean lung sounds for the respiratory support case, suggesting all breathing rate error improvements could be inaccurate. Future work is required to accurately determine if these removed respiratory support sounds are also unintentionally removing these heartbeats/breathing periods, or if there is a subset of recordings in which heart and lung sounds cannot be recovered.

For the computational cost, the proposed NMCF method is not suitable for real-time processing, whereas the proposed NMF method is. There are a couple of ways to improve the computation speed of the proposed NMCF method. Firstly, as the optimal heart and lung sound co-factorisation weights were zero (i.e., no co-factorisation with mixture chest sound recording), a fusion of the NMF and NMCF could be employed. For these zero co-factorisation weights, instead of calculating the associated basis vectors during the separation of the mixture chest sound, instead, similar to NMF, the basis matrices of heart and lung sounds can be pre-computed and fixed. The computation time would be reduced by up to 36% potentially by implementing this. Secondly, the pre-computed NMF basis matrices could be used as the initialisation of the basis matrices for the proposed NMCF method. This initialisation would place the basis matrices closer to the optimal result, meaning fewer iterations of the update algorithm would be required. Thirdly, there is still room for optimisation of the number of reference examples, the number of bases for each sound component, FFT size and hop size, all of which could reduce the size of the basis and activation matrices, overall reducing the computation time. Finally, the multiplicative update algorithm itself could be optimised, for instance using C code MEX functions, as opposed to MATLAB.

An artificial dataset was proposed to quantitatively assess the effectiveness of sound separation with heart, lung, and various noise sounds. However, there are several limitations with this dataset. Firstly, to obtain clean reference sounds, bandpass filtering of 50-250 Hz and 200-1000 Hz was applied to heart and lung sounds. Similarly, for cry noise, highpass filtering of 300 Hz was applied. Whilst these frequency bands contain the majority of the heart, lung, and cry sounds, there can still be prominent sounds outside these frequency bands. Therefore, to more accurately replicate the larger frequency overlap between these sounds and assess the sound separation methods, pure and unfiltered reference sounds would be required. However, this is quite difficult to obtain with newborns. As neonates, in particular preterm neonates, have sporadic breathing patterns of fast periodic breathing followed by a period of no breathing, it may be possible to obtain short segments of clean heart sounds. Obtaining clean reference lung sounds is more difficult as heart sound noise will typically be present. Future collection of chest sounds with digital stethoscope placement on the chest but further away from the heart or on the back may minimise heart sounds and enable clearer lung sounds.

Another option to obtain clean reference sounds is multi-channel blind separation. During data acquisition for the chest sounds, synchronous electrocardiogram for heartbeat detection, chest accelerometer for breathing detection, and secondary microphones for background noise, can provide additional information for more accurate sound separation. The separated sounds can then be used as reference gold-standard sounds for single-channel sound separation methods. Finally, artificially generated sounds with the desired properties could be used to create reference sounds.

The second limitation of the artificial dataset is how cry noise was mixed with lung sounds. As inspiration and expiration cannot occur simultaneously as a newborn is crying, this should be taken into account. For a more appropriate lung sound and cry noise mixture, the periods and order of inspiration and expiration should be taken into account and inserted reasonably.

The third limitation of the artificial dataset is the relatively small evaluation dataset. For respiratory support noise, future collections of pure respiration support noise obtained by placing the stethoscope on the breathing tube will be acquired. Whereas the limitation with heart and lung sounds is not the number of chest sound recordings available, but the difficulty in obtaining clean sounds. Again, potential solutions are artificially creating heart and lung sounds and multi-channel blind separation.

Signal quality estimation of real chest sounds was proposed to quantitatively assess the effectiveness of sound separation on real-world data. This assessment overcomes many of the limitations with artificial datasets and assesses the sound separation methods for the actual data it would be used for. However, while the method attempts to objectively assess heart and lung sound quality, it is not 100% accurate [2], [7]. Therefore, individual comparison of methods with a single recording may not be appropriate due to misclassification errors with the signal quality estimation model. Whereas, overall trends on signal quality improvement and comparison of methods are more appropriate. Additionally, the method assesses the overall quality of the heart and lung sounds. This overall assessment does not directly take into account if significant portions of the heart or lung sound components are also being removed to obtain the “noise-free” sounds. These limitations can also partially explain why the artificial dataset vs real-world dataset sound separation methods differs.
V. CONCLUSION

This paper has reviewed, adapted and tested a wide variety of existing methods for neonatal chest sound separation for clean heart and lung sounds. Additionally, two methods were proposed, namely, NMF and NMCF with reference datasets of heart, lung, and noise sounds. The evaluation results show that our proposed methods outperformed existing methods in both the artificial and real-world datasets. However, further work is still required for sound separation of respiratory support noise, as can be seen in the inferior results compared to existing methods for vital sign estimation. Possible directions are obtaining more clean samples either in the real-world setting or artificially, and looking into deep learning methods.

Overall, this study shows promise in obtaining clean heart and lung sounds in a variety of scenarios. The ability to obtain such high-quality heart and lung sounds enables future studies on the assessment of lung aeration, cardiac murmur detection, prediction of respiratory distress, the effect of treatments on cardio-respiratory conditions, and longitudinal studies on changing chest sounds of neonates as they develop, among other potential studies.

ACKNOWLEDGMENT

E. Grooby thanks and acknowledges the support from the Monash Newborn team at Monash Children’s Hospital, Australia for data collection and Jinyuan He for the analysis of the newborn data in previous works.

REFERENCES

[1] “Newborns: Reducing mortality,” Accessed: Oct. 05, 2019. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/newborns-reducing-mortality
[2] E. Grooby et al., “Neonatal heart and lung sound quality assessment for robust heart and breathing rate estimation for telehealth applications,” IEEE J. Biomed. Health Inform., vol. 25, no. 12, pp. 4255–4266, Dec. 2021.
[3] E. Grooby et al., “Prediction of neonatal respiratory distress in term babies at birth from digital stethoscope recorded chest sounds,” in Proc. 44th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., 2022, pp. 4996–4999, doi: 10.1109/EMBC4829.2022.9871449.
[4] A. Ramanathan et al., “Digital stethoscopes in paediatric medicine,” Acta Paediatrica, vol. 108, no. 5, pp. 814–822, 2019.
[5] L. Zhou, F. Marzbanrad, A. Ramanathan, D. Fattahi, P. Pharande, and A. Malhotra, “Acoustic analysis of neonatal breath sounds using digital stethoscope technology,” Pediatr. Pulmonol., vol. 55, no. 3, pp. 624–630, 2020.
[6] A. Ramanathan et al., “Assessment of breath sounds at birth using digital stethoscope technology,” Eur. J. Pediatrics, vol. 179, pp. 781–789, 2020.
[7] E. Grooby et al., “Real-time multi-level neonatal heart and lung sound quality assessment for telehealth applications,” IEEE Access, vol. 10, 2022, pp. 10934–10948, doi: 10.1109/ACCESS.2022.3144355.
[8] A. King, D. Blank, R. Bhatia, F. Marzbanrad, and A. Malhotra, “Tools to assess lung aeration in neonates with respiratory distress syndrome,” Acta Paediatrica, vol. 109, no. 4, pp. 667–678, 2020.
[9] A. Lahav, “Questionable sound exposure outside of the womb: Frequency analysis of environmental noise in the neonatal intensive care unit,” Acta Paediatrica, vol. 104, no. 1, pp. e14–e19, 2015.
[10] Z. Wang, J. N. da Cruz, and F. Wan, “Adaptive fourier decomposition approach for lung-heart sound separation,” in Proc. IEEE Int. Conf. Comput. Intell. Virtual Environ. Meas. Syst. Appl., 2015, pp. 1–5.
[11] D. B. Springer, L. Tarassenko, and G. D. Clifford, “Logistic regression-HSMM-based heart sound segmentation,” IEEE Trans. Biomed. Eng., vol. 63, no. 4, pp. 822–832, Apr. 2016.
[12] T. Tsalalie and S.ANEI, “Separation of heart sound signal from lung sound signal by adaptive line enhancement,” in Proc. 15th Eur. Signal Process. Conf., 2007, pp. 1231–1235.
[13] A. Mondal, P. Bhattacharya, and G. Saha, “Reduction of heart sound interference from lung sound signals using empirical mode decomposition technique,” J. Med. Eng. Technol., vol. 35, no. 6/7, pp. 344–353, 2011.
[14] M. Pourrazad, Z. Moussavi, and G. Thomas, “Heart sound cancellation from lung sound recordings using time-frequency filtering,” Med. Biol. Eng. Comput., vol. 44, no. 3, pp. 216–225, 2006.
[15] T. H. Falk and W.-Y. Chan, “Modulation filtering for heart and lung sound separation from breath sound recordings,” in Proc. 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., 2008, pp. 1859–1862.
[16] E. Grooby et al., “A new non-negative matrix co-factorisation approach for noisy neonatal chest sound separation,” in Proc. 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., 2021, pp. 5668–5673.
[17] G. Shah, P. Koch, and C. B. Papadias, “On the blind recovery of cardiac and respiratory sounds,” IEEE J. Biomed. Health Inform., vol. 19, no. 1, pp. 151–157, Jan. 2015.
[18] F. Canadas-Quesada, N. Ruiz-Reyes, J. Carabias-Ortu, P. Vera-Candeas, and J. Fuertes-Garcia, “A non-negative matrix factorization approach based on spectro-temporal clustering to extract heart sounds,” Appl. Acoust., vol. 125, pp. 7–19, 2017.
[19] J. Gnitecki, Z. Moussavi, and H. Pasterkamp, “Recursive least squares adaptive noise cancellation filtering for heart sound reduction in lung sounds recordings,” in Proc. 25th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., 2003, pp. 2416–2419.
[20] F. Ghaderi, H. R. Mohseni, and S. Sanei, “Localizing heart sounds in respiratory signals using singular spectrum analysis,” IEEE Trans. Biomed. Eng., vol. 58, no. 12, pp. 3360–3367, Dec. 2011.
[21] A. Mondal, I. Saxena, H. Tang, and P. Banerjee, “A noise reduction technique based on nonlinear kernel function for heart sound analysis,” IEEE J. Biomed. Health Inform., vol. 22, no. 3, pp. 775–784, May 2018.
[22] L. J. Hadjileontiadis and S. M. Panas, “A wavelet-based reduction of heart sound noise from lung sounds,” Int. J. Med. Inform., vol. 52, no. 1-3, pp. 183–190, 1998.
[23] I. Hossain and Z. Moussavi, “An overview of heart-noise reduction of lung sound using wavelet transform based filter,” in Proc. 25th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., 2003, pp. 458–461.
[24] K.-H. Tsai et al., “Blind monaural source separation on heart and lung sounds based on periodic-coded deep autoencoder,” IEEE J. Biomed. Health Inform., vol. 24, no. 11, pp. 3203–3214, Nov. 2020.
[25] B. Gao, W. L. Woo, and S. S. Dlay, “Unsupervised single-channel separation of nonstationary signals using gammatone filterbank and itakura–saito nonnegative matrix two-dimensional factorizations,” IEEE Trans. Circuits Syst. I: Regular Papers, vol. 60, no. 6, pp. 662–675, Mar. 2013.
[26] J. Le Roux, F. J. Weninger, and J. R. Hershey, “Sparse NMF–half-baked or well done?” Mitsubishi Electric Research Labs (MERL), Cambridge, MA, USA, Tech. Rep. TR2015-023, vol. 11, pp. 13–15, 2015.
[27] R. Nersisson and M. M. Noel, “Heart sound and lung sound separation algorithms: A review,” J. Med. Eng. Technol., vol. 41, no. 1, pp. 13–21, 2017.
[28] H. W. Chong, L. Cai Cai, A. Lin, and R. J. Skillington, “Electronic stethoscope,” U.S. patent application publication US 2015/0190110 A1, 2015.
[29] “VoiceRecorder and Tapmedia Ltd, London, United Kingdom,” Accessed: Dec. 31, 2019. [Online]. Available: http://www.tapmedia.co.uk
[30] E. Vincent, R. Griboval, and C. Févotte, “Performance measurement in blind audio source separation,” IEEE Trans. Audio, Speech, Lang. Process., vol. 14, no. 4, pp. 1462–1469, Jul. 2006.
[31] C. Févotte, R. Griboval, and E. Vincent, “BSS_EVAL toolbox user guide–Revision 2.0,” Tech. Rep., pp. 19, fnmria-00564760, 2005.
[32] J. Le Roux, S. Wisdom, H. Erdogan, and J. R. Hershey, “SDR–half-baked or well done?” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process., 2019, pp. 626–630.
[33] D. Jain and E. Bancalari, “Neonatal monitoring during delivery room emergencies,” Seminars in Fetal and Neonatal Medicine, vol. 24, no. 6, 2019, Art. no. 101040.