IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 22, 2021, accepted January 15, 2022, date of publication January 18, 2022, date of current version January 31, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3144355

Real-Time Multi-Level Neonatal Heart and Lung Sound Quality Assessment for Telehealth Applications

E. GROOBY[1,2], (Member, IEEE), C. SITAULA[1], D. FATTAHI[1,3], R. SAMENI[4], (Senior Member, IEEE), K. TAN[5], L. ZHOU[5], A. KING[5], A. RAMANATHAN[5], A. MALHOTRA[5], G. A. DUMONT[2], (Life Fellow, IEEE), AND F. MARZBANRAD[1], (Senior Member, IEEE)

[1]Department of Electrical and Computer Systems Engineering, Monash University, Melbourne, VIC 3800, Australia
[2]Electrical and Computer Engineering Department, The University of British Columbia, Vancouver, BC V6T 1Z4, Canada
[3]School of Electrical and Computer Engineering, Shiraz University, Shiraz 71348-51154, Iran
[4]Department of Biomedical Informatics, Emory University, Atlanta, GA 30322, USA
[5]Paediatrics Department, Monash University, Melbourne, VIC 3800, Australia

Corresponding author: E. Grooby (ethan.grooby@monash.edu)

This work was supported by the Monash Institute of Medical Engineering (MIME). The work of E. Grooby was supported by the MIME-Monash Partners-The Commonwealth Scientific and Industrial Research Organisation (CSIRO) sponsored the Ph.D. Research Support Program and the Research Training Program (RTP). The work of A. Malhotra was supported in part by the Kathleen Tinsley Trust, and in part by the Cerebral Palsy Alliance Research Grant. The work of F. Marzbanrad was supported by the Advancing Women's Research Success Grant Program.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Monash Health Human Research Ethics Committee under Application No. HREA/18/MonH/471.

ABSTRACT In this study, a new method is proposed to assess heart and lung signal quality objectively and automatically on a 5-level scale in real-time, and to assess the effect of signal quality on vital sign estimation. A total of 207 10 s long chest sounds were taken from 119 preterm and full-term babies. Thirty of the recordings from ten subjects were obtained with synchronous vital signs from the Neonatal Intensive Care Unit (NICU). As a reference, seven annotators independently assessed the signal quality. For automatic quality classification, 400 features were extracted from the chest sounds. After feature ranking and selection, class balancing, and hyperparameter optimization, a variety of multi-class and ordinal classification and regression algorithms were trained. Then, heart rate and breathing rate were automatically estimated from the chest sounds. For the deep learning model, YAMNet, a deep convolutional neural network pre-trained on the AudioSet-Youtube corpus for sound classification was used. After modification of the final output layers of the neural network and class balancing, transfer learning was applied to YAMNet for heart and lung signal quality classification. The results of subject-wise leave-one-out cross-validation show that the best-performing models had a balanced accuracy of 56.8% and 51.2% for heart and lung qualities, respectively. The best-performing models for real-time analysis (<200 ms) had a balanced accuracy of 56.7% and 46.3%, respectively. Our experimental results underscore that increasing the signal quality leads to a reduction in vital sign error.

INDEX TERMS Breath sound, deep learning, heart rate, heart sound, neonatal monitoring, ordinal regression, phonocardiogram (PCG), signal quality assessment, respiration rate, telehealth.

I. INTRODUCTION

The neonatal period is the most vulnerable time for survival, with 1.7% of live births resulting in mortality, totalling 2.4 million worldwide, in 2019 alone [1]. To address this major issue, the United Nations created the 3.2.2 Sustainable Development Goal, to reduce neonatal mortality to 1.2% of live births by 2030 [2].

Stethoscope-record chest sounds contain important cardiac and respiratory information that inform neonatal health

The associate editor coordinating the review of this manuscript and approving it for publication was Easter Selvan Suviseshamuthu.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

status. This information can enable timely assessment for signs of serious health risks to potentially improve neonatal survival and reduce long-term morbidity risk [3]–[5].

However, low-quality chest sounds, due to the noise from either external environment, other internal body sounds, or the device itself, hinder the usage of stethoscopes. Low-quality recordings complicate monitoring and diagnosis, or at worse lead to misdiagnosis [6], [7]. Whilst having low-quality chest sounds is unavoidable, identification and exclusion of low-quality recordings help to improve remote monitoring. Real-time automated quality assessment of heart and lung sounds would address this gap by assisting the users in obtaining better diagnostic-quality recordings and ensuring the reliability of diagnosis.

Previous research on heart signal quality analysis has mainly focused on the binary classification of heart sounds into high and low-quality on adult populations. In our past work, these methods were reviewed in detail, adapted and expanded for the neonatal population [6]. To summarise, heart sound recordings were represented in several ways: time and frequency domain, autocorrelation signal, wavelet decomposed signal and segmented heart signal into S1 and S2 sounds [8]–[12]. Features were then extracted from these representations including statistical features (variance, skewness and kurtosis), predictive fitting coefficients, segmentation quality and agreement, Mel-frequency coefficients (MFCC), entropy and power. These features were then used to develop a dynamic classifier with 96% specificity, 81% sensitivity and 93% accuracy. These results were shown to be superior to the individual implementation of past heart signal quality estimation techniques [6].

To date, limited studies have investigated lung sound quality assessment, either relying on an external reference signal or by generating an artificial set of low and high-quality lung sounds [13], [14]. In our past work, heart sound quality methods were adapted for lung sound quality analysis, with 86% specificity, 69% sensitivity, and 82% accuracy, for binary classification of low vs high-quality [6].

There are several key limitations with past works in heart and lung signal quality classification. Firstly, except for Kala *et al.* [13], who created a model to estimate signal to noise ratio in artificial mixtures of clean lung sounds with background noise, the majority of works including our previous work, only provide a binary assessment of signal quality as either high or low quality. As typically the original signal quality annotations are on a 5-point scale or continuous metric, threshold selection to determine a binary high or low quality can be quite arbitrary. In particular, recordings around the boundary of high and low quality are difficult to classify, and the binary classification is not highly informative. Instead, through providing a finer scale of signal quality assessment, representative of the original annotations, users can make more informed decisions on the diagnostic quality of the recordings. Additionally, a relevant and important research area of interest is denoising and sound separation to obtain clean heart and lung sounds from noisy chest sound recordings [15]. The assessment of these methods is difficult in the neonatal context, as obtaining high-quality heart and lung sounds to create artificial mixture test sets is very difficult. Even if artificial mixture test sets are generated, they only partially test the effectiveness of these methods. Instead, a continuous fine-scale signal quality assessment method is required, to accurately assess these denoising, and heart and lung sound separation methods on real-world neonatal chest sounds.

The second limitation with past works including our own, is the lack of assessment of the appropriateness of these methods to be applied in the telehealth context. For heart and lung signal quality classification to adequately inform users of obtaining better diagnostic-quality recordings, real-time feedback is required. Thus, an assessment of the computation cost of the signal quality classification models is required.

The third limitation is the lack of assessment of the signal quality definitions on the accuracy of extracting vital signs. Since one of the goals of obtaining high-quality signals is to achieve an accurate heart and breathing rate, it should be verified that this is captured in the definition of signal quality and what this relationship is. This was partially addressed in our past work with manual annotations for heart and breathing rate [6]. However, to make more accurate conclusions, vital signs obtained from a gold-standard reference electrocardiogram should be used.

The final limitation is the requirement of hand-crafted features to estimate heart and lung signal quality. Hand-crafted features generally reduce the generalisability of the model. For instance, existing works features typically rely on adult-based heart and lung parameters, which have to be modified to be applicable for the neonatal population [6]. Furthermore, due to less research in lung signal quality analysis, there are fewer hand-crafted features, resulting in inferior results, in comparison to heart signal quality analysis [6]. The development of deep learning models, which take the time-frequency representation of the audio chest sound signal, offers the potential for removing the need for hand-crafted features for the estimation of signal quality [16].

The key contribution of this research is the automated real-time multi-level quality rating of neonatal chest sounds, which can guide the user during auscultation and sound recording. The development of a real-time system enables an objective assessment of signal quality for the user to obtain better diagnostic quality recordings. Another contribution of this research is to extend our previous feature-based binary signal quality classification model to a five-level quality scale and compare this to the deep learning model, YAMNet. This is achieved by introducing new features, providing a more detailed signal quality assessment for the feature-based model, and applying transfer learning to a pre-trained convolutional neural network for the deep learning model. Through providing a finer scale signal quality assessment, users can make more informed decisions on the diagnostic quality of the recordings. Unlike previous studies, the processing time to extract features is assessed to determine the applicability

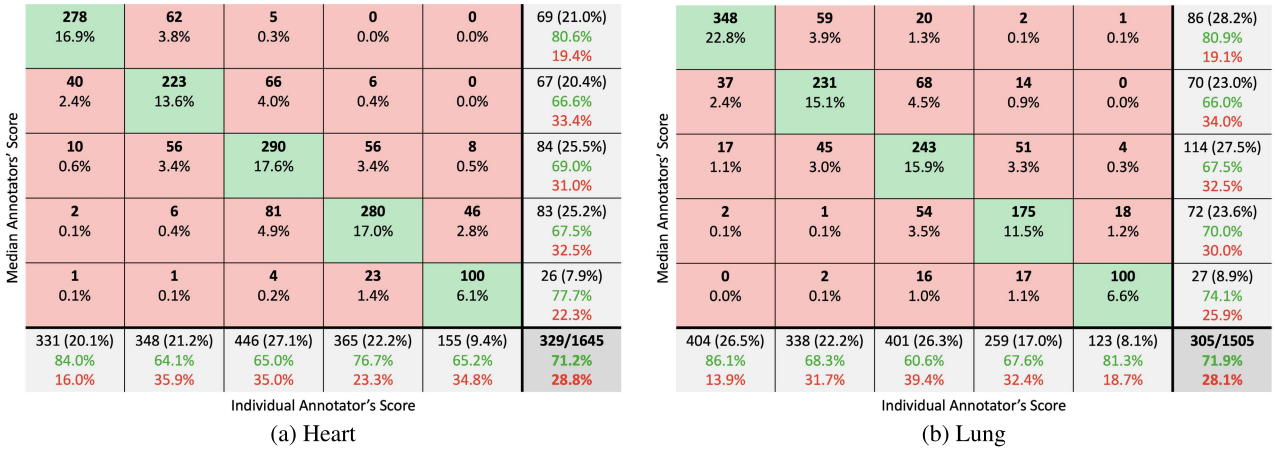


FIGURE 1. Confusion matrix of overall annotators' scores for a) heart signal quality and b) lung signal quality. The x-axis represents the individual annotator's score and the y-axis median annotators' score for each chest sound recording. Green and red squares show the number of annotations that have the same or different overall median annotation respectively. Additionally, the total percentage of annotations in each square is shown. Grey squares at the end of each row show total recordings for a particular median annotator score, with the percentage of total annotations in brackets, and percentage agreement and disagreement in that row shown in green and red, respectively. Grey squares at the end of each column show similar information, except for total annotations for each individual annotator's score. Dark grey square summarises the information of the total number of recordings/total number of annotations and the overall percentage of annotations agreeing with median annotation.

of heart and lung sound quality assessment in real-time. Since one of the key purposes of acquiring high-quality sounds is to obtain accurate vital sign estimates, the relationship between the heart and lung sound quality with heart and breathing rate accuracy is also assessed using gold standard NICU recordings.

The rest of this paper is organized as follows. Section II presents details of the proposed signal quality assessment models. Evaluation and results are presented in Section V. Feasibility of real-time analysis and comparison of signal quality with heart and breathing rate error with the models are discussed in Sections IV and III, followed by a discussion in Section VI. Section VII concludes the work, with future perspectives.

II. METHODS

A. DATA ACQUISITION AND PREPROCESSING

The study was conducted at Monash Newborn, Monash Children's Hospital. It was approved by the Monash Health Human Research Ethics Committee (HREA/18/MonH/471). A total of 318, 60 s recordings from the right anterior chest of preterm and term newborns were obtained using a digital stethoscope [17], [18]. Synchronous vital signs based on electrocardiogram for reference heart and breathing rate were collected for 32 recordings, as further detailed in Section III. The chest sounds were low-pass filtered to avoid aliasing and down-sampled to 4 kHz. Recordings significantly damaged from artifacts making lung and heart sounds impossible to recover were automatically removed using methods presented in our previous work [6]. Next, 10 s segments containing heart, breathing or both sounds were manually extracted. After excluding the invalid recordings, a total number of 207 signals (119 subjects) remained, 30 (10 subjects) of which had synchronous vital signs. These 30 recordings were held out only for testing the trained models.

B. ANNOTATION SETS AND QUALITY ANNOTATIONS

Randomized heart and lung pools were created from 207 raw recordings plus 207 frequency filtered recordings, resulting in 414 heart and 414 lung recording pools [6]. Annotated signal quality was used as ground truth, provided by 3 clinicians and 4 electrical engineers familiar with biomedical auscultation, producing a set of 5 annotations per recording. To have a more reliable training set, recordings with inter-rater agreement less than or equal to 0.2 based on Fleiss kappa, were removed [19]. This resulted in a total of 329 heart recordings and 305 lung recordings. The short-listed heart and lung recordings resulted in an inter-rater agreement of 0.37 and 0.39, respectively, which correspond to a fair agreement. Median annotators signal quality was then used to determine the signal quality for each recording. The resultant signal quality distributions are shown in Figure 1.

Further details of the methodology can be found in our previous work [6].

C. FEATURE-BASED QUALITY ASSESSMENT MODEL

1) FEATURES

From our recent work, a total of 182 and 187 features for lung and heart sound quality classification are extracted [6]. As strong high-quality heart sound can act as noise and reduce the quality of lung sound and vice versa, the heart and lung features were combined together and used for both heart and lung signal quality classification.

The initial feature set was also expanded in 3 ways, to overall create a total of 254 features. Firstly, both 5 s truncated and full autocorrelation signal of Hilbert envelope was used to calculate the autocorrelation-based features, as proposed by Springer *et al.* [8]. Secondly, Hilbert, homomorphic, Shannon, Short-Time Fourier Transform (STFT), power for 40-60 Hz and 3rd-level detailed wavelet coefficients with *rbio3.9* wavelet envelopes were calculated, as they

TABLE 1. Feature set used for automatic classification of heart and lung sounds.

Number	Title	Description
1	Audio Sample Entropy	Signal complexity measure, which is high for unpredictable signals. The similarity between two epochs is determined at a threshold r for M time points. This is calculated on the down-sampled 30 Hz signal with M=2 and r=0.1 [10].
2	Percentage Clipping	Clipping is an undesirable form of noise due to the recorded signal exceeding the maximum limit of the digital stethoscope. The normalised magnitude of the audio signal (range [0,1]) is calculated, and the percentage of points above 0.97 is determined [14].
3-4	Mean Rate Average Energy	Recording was down-sampled to either 1000 Hz or 2000 Hz. The average temporal energy variation along each frequency channel of the power spectral density of the recording was calculated over the range 2-32 Hz [13].
5-6	Percentage Heart Contamination	Percentage of the recording with prominent heart sounds was determined. This was achieved with 50-250 Hz 4^{th} order Butterworth bandpass filtering of recording. Then wavelet decomposition was performed with Symlet wavelet at depth 3. The 3 approximation coefficients were then normalized and multiplied together to get a representation of prominent heart sound peaks. Percentage of this signal exceeding 0.1 and 0.2 were then calculated [14].
7	High Frequency Variance	Audio signal was 2^{nd} order high pass filtered at 700 Hz and variance calculated [20], [21].
8-18	LPC	10^{th} order linear predictive coefficients (LPC) were calculated [22].
19-21	Entropy	Shannon, Renyi and Tsallis entropy of audio recording is calculated [22].
22-23	Degree of Periodicity	Heart sounds and to an extent breathing sounds can be considered quasiperiodic. The degree to which heart and breathing sounds are periodic in recording are calculated in the ranges 70-220 bpm and 15-80 bpm respectively [6], [23], [24].
24	Autocorrelation Kurtosis	Kurtosis of autocorrelation signal is calculated. Kurtosis is a measurement of the degree to which the probability distribution function clusters at the tails [10].
25-26	Autocorrelation Sample Entropy	Sample entropy with M=2 and r=0.2 is calculated on the 30 Hz down-sampled full and 5 s truncated autocorrelation signal [6], [23].
27-28	Autocorrelation Cycle Duration	Cycle duration is calculated based on the peak in the autocorrelation signal corresponding to 70-220 bpm and 15-80bpm for heart and lung respectively [23].
29	Cry Power	Based on previous work, the power ratio in the 295-406 Hz frequency region, in which crying can be most easily identified is calculated [6].
30-42	Power	For 2000 Hz down-sampled signal, power ratio is determined for 0-100 Hz, 100-200 Hz, 200-300 Hz, 300-400 Hz, 400-500 Hz, 500-600 Hz, 600-700 Hz, 800-900 Hz and 900-1000 Hz frequency ranges [22]. Similarly, this was done for frequency ranges 24-144 Hz,144-200 Hz and 200-1000 Hz [23].
43	Power Centroid	For 2000 Hz down-sampled signal, power spectral density is calculated and centroid calculated [22].
44-51	Linear Dependency of PSD	The time-frequency power spectral density (PSD) is calculated on the original and 2000 Hz down-sampled signal. The frequency component is then compressed to 15 evenly spaced bins 1-15. Singular value decomposition is then calculated on bins 1-5, 6-10 and 11-15, and the ratio of 2^{nd} and 1^{st} components is determined [25].
52-61	Wavelet Entropy	Wavelet decomposition is performed at depth 5 using 4^{th} order Daubechies wavelet. Shannon, Tsallis and Renyi entropies are then calculated on the 5^{th} level approximation coefficients and 3^{rd}, 4^{th} and 5^{th} level detailed coefficients. Additionally, log-variance is calculated on the 3^{rd} level detailed coefficients [22].
62-63	Wavelet RMSSD and ZCR	Wavelet decomposition is performed at depth 2 using 8^{th} order Daubechies wavelet. Zero crossing rate (ZCR) and root mean square of successive differences (RMSSD) are calculated on 2^{nd} level approximation coefficients [12].
64-65	Wavelet ZCR	Recording was down-sampled to 1000 Hz and high pass filtered at 20 Hz. Wavelet decomposition is then performed at depth 1 using 2^{nd} order Daubechies wavelet. After normalization, peaks were then detected [26]. ZCR was calculated with a threshold of 85^{th} percentile of wavelet decomposed signal and 58^{th} percentile of detected peaks values [11].
66-82	MFCC	MFCC calculated with a window length of 25ms, overlap length of 15ms and with 13 coefficients plus log energy. The minimum, maximum and skew are then calculated for all 14 signals and then averaged [22].
83-84	Fundamental Frequency	Fundamental frequency calculated using the cepstral method with window length 25ms and overlap 15ms, in the 50-1000 Hz range. The percentage of frames with a fundamental frequency less than 250 Hz is calculated as this corresponds to newborn crying [14]. Additionally, the overall fundamental frequency is determined by plotting the values on a histogram and taking the value of the largest bin [10].
85-100	Envelope Sample Entropy	All envelopes were down-sampled to 30 Hz and sample entropy with M=2 and r=0.2 was determined [23].
101-107	Envelope Variance	Variance of all heart-based envelopes were calculated [23].
108-121	Envelope Heart Cycles	For heart-based envelopes, using the estimated cycle duration from the autocorrelation signal, the envelope is divided up into the same length segments, and the correlation between these segments is calculated. The average and standard deviation of the correlation values is then determined [23].
122-128	Envelope Heart Rate Variability	Using a sliding window of 3 s, heart rate is calculated from the autocorrelation of the heart-based envelopes. The average heart rate and heart rate variability are then reported [23].
129-138	Percentage Bad Segmentation	Recordings are segmented using either the method proposed by Schmidt et al. or Springer et al. [27], [28]. Both methods used duration-dependent hidden Markov models to separate the recording into four different states, namely, S1, systole, S2 and diastole. Segments assigned the same state are grouped, and outliers corresponding to poor segmentation are identified [20].
139-140	Segmentation Quality	Recordings are segmented using either the method proposed by Schmidt et al. or Springer et al. [27], [28]. For each S1 and S2 segment, 4 level MFCC decomposition is performed, which is re-sampled to length 14 and transformed into 70 (4 coefficients + log energy $\times$ 14) length cepstral vector. Total cepstral distances between all S1 segments (dS1) and all S2 segments are calculated (dS2). Segment quality is then represented as the combined total of cepstral distances (dS1 + dS2) [29].
141-146	Percentage Abnormal Segmentation	For Schmidt et al. and Springer et al. segmented signals, ZCR, RMSSD and SD1 of Poincaré plot are calculated for the systolic and diastolic segments separately, and the systolic and diastolic segments combined. The difference between the systolic and diastolic segment values divided by the combined segment is calculated. The percentage of segments above 0.8, 0.8 and 0.6 for RMSSD, SD1 and ZCR respectively are then calculated [11].

commonly represent heart signals [23], [27], [30], [31]. Similarly, log-variance, variance fractal dimension, spectral energy and powers in the 0-500 Hz, 150-300 Hz, 300-450 Hz and 150-450 Hz bands were calculated as they are commonly used to represent lung signals [32]–[35]. As opposed to just using Hilbert envelope as in the past work, all heart signal based envelopes were used to calculate hidden semi-Markov model (HSMM) quality features [9]. Finally, the percentage acceptable windows feature was modified from our past work [6]. The original feature used a sliding window of 2200 ms with 25% overlap, and calculated the number of heart peaks within each window using a method proposed by Gieraltowski *et al.* [26]. The percentage of windows containing the normal range of 2-4 heart peaks was then calculated and used to estimate signal quality [11], [26]. Percentage windows with the number of peaks 4-7 and 2-8 were also considered, as these correspond to approximately the 5^{th} and 95^{th} percentile heart rate, and the full heart rate ranges of newborns, respectively [6], [36], [37]. Additionally, Springer *et al.*, Schmidt *et al.*, and Liang *et al.* methods were used to detect S1 and S2 heart peaks, as opposed to just heart peak detection using the Gieraltowski *et al.* method previously. The percentage windows in acceptable ranges of 4-8 (original), 9-14 (5^{th}-95^{th} percentile), and 5-16 (full range) were then determined based on these detected S1 and S2 peaks [6], [11], [27], [28], [31], [36], [37]. The percentage acceptable windows features was also adapted for lung with 4 s sliding window with 25% overlap and peak detection of inspiration and expiration peaks using methods developed in our past work [6]. Percentage of windows with 1-4 and 1-5 corresponding to 5^{th} and 95^{th} and the full range of respiratory rate, respectively, were then calculated [6], [36], [37].

146 additional features based on previous literature were also extracted, as summarized in Table 1. Features 85-100 used lung-based envelopes and features 85-128 used the aforementioned heart-based envelopes. In total 400 features were extracted for lung and heart sound quality classification. The source codes for these features are provided online in [38].

2) FEATURE RANKING AND SELECTION

The training set was normalized to have zero means and unit variance, with these same scaling and shifting values used on the test set.

For feature ranking, the training set was class balanced with random up-sampling with replacement and maximum Relevance Minimum Redundancy (mRMR) algorithm with Mutual Information Difference (MID) method used [39]. The mRMR algorithm maximizes relevance D (Equation 1) and minimizes redundancy R (Equation 2) based on their difference (Equation 3), in a first-order incremental search to rank the most important features as calculated below:

$$\max D(S, c),\ D = \frac{1}{|S|} \sum_{x_i \in S} I(x_i; c) \tag{1}$$

$$\min R(S),\ R = \frac{1}{|S|^2} \sum_{x_i, x_j \in S} I(x_i, x_j), \tag{2}$$

$$\max \phi(D, R),\ \phi = D - R \tag{3}$$

where S is the feature set, x_i, x_j are individual features, I is mutual information, and c is the target class.

The mean-square error was plotted against the number of features used based on the mRMR algorithm in Figure 5. From this figure, heart classification performance plateaus from feature 5 onwards and lung classifier performance degrade after feature 20. To find the region of best performance and minimize overfitting, the ranges of the top 5-15 features for heart and top 5-20 features for lung were selected for hyperparameter optimization. Note, while the same overall 400 feature set was used for both heart and lung signal quality classification, after feature ranking and selection, the top features used are specific to each classifier, as shown in Figures 2 and 3.

3) CLASSIFICATION

The overall model is shown in Algorithm 1, which takes in all recording features, patient assignment, signal quality annotations and hyperparameters as input to train the classifier.

As shown in Table 1, the distribution of signal qualities is not even. In particular, there are few recordings of high-quality, this is because recording in a neonatal intensive care environment is challenging with a large range of noises occurring. To resolve this, patient-wise class balancing was performed with the minority class being randomly upsampled with replacement.

Two groups of classifiers were implemented. The first group was standard regression methods that either had no parameters (ordinary least squares regression, AdaBoost, gradient boosting, bagging and random forest), had regularisation strength optimised through 5-fold cross-validation (ridge regression [alpha = 0.1, 0.5, 1.0, 5, 10.0, 50, 100, 500, 1000], LASSO, Elastic-Net [l_1 ratio = 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 0.7, 0.9, 0.95, 0.99, 0.995, 0.999, 1.0], least angle regression, LASSO with least angle regression [max iterations = 50] and orthogonal matching pursuit) or had numerous parameters optimised using 5-fold cross-validation grid search parameter optimisation based on mean square error (support vector machine, decision tree and k-nearest neighbours) [40]. The second group was ordinal regression methods with either single parameter (least absolute deviation [max iterations = 5000]) or had numerous parameters optimised using grid search (logistic model with all or immediate threshold, ridge and support vector machine) [41].

With the standard regression method group, test set outputs were restricted to be in the range 1-5, whereas the ordinal regression is similar to multi-class classification, except that the order of the annotations is factored into the training. That is, if the correct output is 1, then a mis-classification of 2 is better than a mis-classification of 3. In fact, existing multi-class classifiers can be modified to be ordinal regression

Algorithm 1 Quality Assessment Model

Require: *features*, *patientList*, *annotations*, *params*

for each *patient* in *patientList* **do**
  testSet ← *features*(*patient*)
  testLabel ← *annotations*(*patient*)
  trainSet ← *features*(!*patient*)
  trainLabel ← *annotations*(!*patient*)
  trainPatients ← *patientList*(!*patient*)
end for
scaler ←StandardScaler().fit(*trainSet*)
trainSet ← *scaler*.*transform*(*trainSet*)
testSet ← *scaler*.*transform*(*testSet*)
trainSet_balanced, *trainLabel_balanced* ← RandomOverSampler(*trainSet*, *trainLabel*)
topFeatures ← mRMR(*trainSet_balanced*, *trainLabel_balanced*, 'MID')
folds ← StratifiedCV(*trainSet*, *trainLabel*,*trainPatients*,splits=5)
for each *fold* in *folds* **do**
  fold ← RandomOverSampler(*fold*)
  *R*1 ← GridSearch(*Regressor*(), *params*, *mse*)
  *R*1.*fit*(*fold*)
  *R*2 ← RegressorCV(*params*)
  *R*2.*fit*(*fold*)
end for

classifiers, which was done for the Support Vector Machine classifier [42].

Patient-wise cross-validation was performed and all parameters tested in the grid search hyperparameter are listed below.

- Support Vector Machine (SVM):
  - Kernel = Radial basis function or linear kernel
  - Kernel Coefficient = 0.1, 0.01, 0.001, 0.0001, inverse of number of features, or inverse of number of features times variance
  - Regularization parameter C = 0.02, 0.04, 0.08, 0.16, 0.32, 0.64, 1.28, 2.56 or 5.12
- Decision Tree (Tree):
  - Measure quality of tree split = Mean square error, Friedman mean square error, mean absolute error, Poisson deviance
  - Max depth of tree = Any, 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10
  - Max features in each split = All features, square root of all features, or log2 of all features
- K-Nearest Neighbours (KNN):
  - Number of neighbours = 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10
  - Weight function = Uniform or inverse of the distance
  - Algorithm to compute nearest neighbours = Brute-force search, BallTree or KDTree
  - Definition of distance = Manhattan distance (l1) or euclidean distance (l2)
- Logistic Model with All or Immediate Threshold:
  - Alpha = 0, 0.1, 0.5, 1, 5, 10, 50, 100, 500, 1000
  - Max iterations = 10,000
- Ordinal Ridge:
  - Alpha = 0, 0.1, 0.5, 1, 5, 10, 50, 100, 500, 1000

D. DEEP LEARNING-BASED QUALITY ASSESSMENT MODEL

Deep learning models, which have already been used in several problems such as neonatal bowel sound classification [43], offer the potential to remove the need for hand-crafted features. However, successfully training such a model typically requires labelled audio examples in the order of 100,000s to millions, equating to thousands of hours of recordings [44]. Chakraborty *et al.* [16] created a deep learning convolutional neural network for binary heart signal quality classification. The model took as input the time-frequency representation of the heart sound recording, as calculated using the STFT. The dataset used in that study included 3,240 recordings ranging from 5 s to over 120 s in duration. As this large dataset does not include neonatal recordings like ours, it may not be able to extract appropriate features from our dataset. Given this, direct training of a deep learning model is not feasible for heart and lung quality classification.

To address this issue of inadequate data, transfer learning is applied. Transfer learning utilises a deep neural network that has already been pre-trained on a related task. This pre-training enables useful weights and relevant abstract features in the early layers of the neural network that are relevant to our task to be obtained. The final layers of the pre-trained neural network are then modified to suit the new task of heart and lung signal quality classification. Then, the modified neural network is trained on the relatively small dataset of neonatal heart and lung signal quality labels.

1) TRANSFER LEARNING

YAMNet, a pre-trained deep convolutional neural network model, is considered in this study [44]. YAMNet predicts 521 audio event classes based on the AudioSet-YouTube corpus [45], [46]. These classes classify a large variety of sound events, including heart and respiratory sounds, and relevant noise sources such as human sounds and generic background noises [45], [46]. Given heart and lung sound quality is relevant to identifying heart, lung and noise sounds, this is a very applicable model for transfer learning.

YAMNet is based on the MobileNet architecture, whereby depth-wise separable convolutions instead of standard convolutions are used [44], [47]. This method of convolutions drastically decreases the computational cost of the model, making it suitable for real-time processing on a mobile phone [47]. As an input, YAMNet initially takes a 0.98 s audio segment sampled at 16 kHz. Mel spectrogram is then calculated with a window length of 25 ms, an overlap of length 15 ms, and 64 frequency bands in the range of 125-7500 Hz. Mel spectrogram is a 96 × 64 image that is inputted into the YAMNet

model [44]. As our audio recordings are of 10 s length and sampling frequency of 4 kHz, these recordings were upsampled to 16 kHz and segmented with a window length of 0.98 s and 50% overlap. Overall, this generates 19 segments per 10 s recording.

For transfer learning of YAMNet, the final three layers are removed, namely, the fully connected, softmax and classification layers. A new fully connected layer and regression layer are then added to the model. The new fully connected layer input is a $1 \times 1 \times 1024$ image, and the output of the new final regression layer is a single value representing the signal quality of the recording segment.

2) CLASSIFICATION

Due to the high computational cost of training the model, patient-wise hyperparameter optimisation was not possible. Instead, patient-wise 5-fold cross-validation was performed to identify the optimal parameters for training the model. In each fold, class balancing was performed with the minority class being randomly upsampled with replacement. Then, the modified YAMNet model was trained with Adam stochastic optimizer with a mini-batch size of 128 audio segments and shuffling of the training set every epoch [48]. The two hyperparameters that were optimised are the max epochs (10, 20 or 30) and the number of early layers that have their weights frozen (0,5,10,...,80). The rationale of optimising these two parameters is to avoid overfitting the dataset and maintain the early-layer features based on the sound classification that are relevant to signal quality classification.

It was found that max epochs of 10 with 40 layers and 60 layers' having their weights frozen, produced the best results, for heart and lung signal quality classification respectively. Using these fixed parameters, patient-wise cross-validation was then performed.

III. HEART RATE AND BREATHING RATE ERROR

The purpose of this section is to analyse the relationship between heart signal quality and heart rate estimation error, and similarly for lung signal quality and breathing rate estimation error.

Using the 30 audio recordings with synchronous electrocardiogram, heart rate and breathing rate were automatically calculated every second with the inbuilt Dräger Infinity® M540 system algorithm [49]. The electrocardiogram is considered the gold standard method for estimation of heart and breathing rate and is used as a reference [4]. These recordings were not involved in training the regression model for quality assessment and were held out for testing.

For the heart audio recordings, heart rate in beats per minute was estimated every second with a sliding window of 3 s. A sliding window of 3 s was chosen as this is a sufficient length to obtain a minimum of three heartbeats, necessary for accurate heart rate estimation. Two methods were used to estimate heart rate. Firstly, using the method proposed by Schmidt *et al.* [28], where the autocorrelation of the Hilbert Envelope is calculated. The maximum peak is then detected in the autocorrelation signal between the bounds of 70-220 beats per minute. The range of 70-220 beats per minute is chosen as this is the typical heart rate range for newborns [6], [50], [51]. The second method proposed by Springer *et al.* [27], uses the initial estimate of heart rate from the Schmidt *et al.* method as input into a duration-dependent hidden Markov model, to segment the heartbeats into four states, namely S1, S2, systolic and diastolic.

For lung audio recordings, breathing rate in breaths per minute was estimated every second with a sliding window of 6 s. Similarly as before, a sliding window of 6 s was chosen as this is a sufficient length to obtain a minimum of three breathing periods, which is necessary for accurate breathing rate estimation. For breathing rate estimation, power spectral envelope is calculated for the frequency range 300-450 Hz, and then peak detection is performed [6].

Using the feature-based quality assessment model proposed in Section II-C, this was trained using the heart and lung recordings that did not have synchronous electrocardiogram recordings. Heart and lung signal quality for the 30 synchronous recordings was then estimated using this trained model.

IV. REAL-TIME PROCESSING

The top 20 features based on the mRMR algorithm are shown in Figure 2. The median time for feature extraction was calculated using MATLAB 2021a with MacBook Pro CPU 2.3 GHz 8-Core Intel i9. For extracting all 20 features, 1.46 s and 2.16 s is required for heart and lung, respectively. The best performing classifiers used 15 and 19 features for heart and lung, which corresponded to 1.12 s and 2.16 s, respectively.

Time-consuming features to calculate were:

- Sample entropy of autocorrelation signal which takes 850ms and 210ms for full and 5 s truncated autocorrelation signal.
- Heart segmentation based features as Schmidt *et al.* and Springer *et al.* segmentation take 120 ms and 80 ms, respectively [27], [28]
- Heart and lung-based singular value decomposition, which take between 50-250 ms per feature
- STFT envelope-based features, as STFT envelope takes 120 ms to calculate
- Mean rate average energy at 1000 Hz and 2000 Hz sampling frequency, that take 640 ms and 1.28 s to calculate

All features above were removed, except for Springer *et al.* segmentation, as many of the top features in both heart and lung utilised segmentation based features. Figure 3 shows the new top 20 features, which take 160 ms and 200 ms to extract for heart and lung, respectively. The best performing classifiers used 13 and 14 features for heart and lung, which corresponded to 130 ms and 120 ms, respectively.

For the deep learning model, both heart and lung sound classification take the same time. To represent the 10 s audio in the Mel time-frequency domain takes less than 10 ms. For

TABLE 2. Heart rate and breathing rate errors.

	Mean Absolute Error (bpm)*					% Acceptable**				
Signal Quality	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
Heart Rate Schmidt et al. [28]	44.2	18.2	12.3	5.2	2.3	26.8	38.5	70.3	74.5	92.6
Heart Rate Springer et al. [27]	49.3	20.9	13.8	7.5	4.9	9.8	15.4	40.7	49.0	66.7
Breathing Rate [6]	30.6	14.1	12.0	4.6	2.8	0	9.1	31.2	66.7	66.7

*bpm denotes beats or breathing periods per minute. Mean absolute error is calculated based on the heart/breathing rate estimation method in comparison to gold standard synchronous electrocardiogram estimation.
**Acceptable % refers to the proportion of the recordings in which heart/breathing rate error is less than 5 bpm

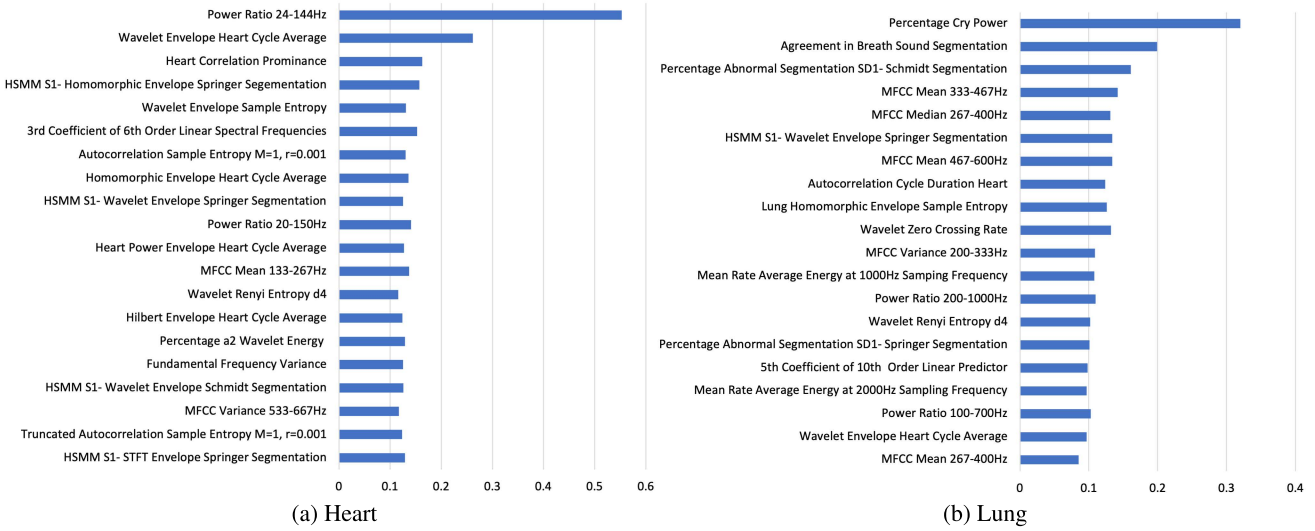


FIGURE 2. The top 20 features based on the mRMR algorithm with the MID method [39].

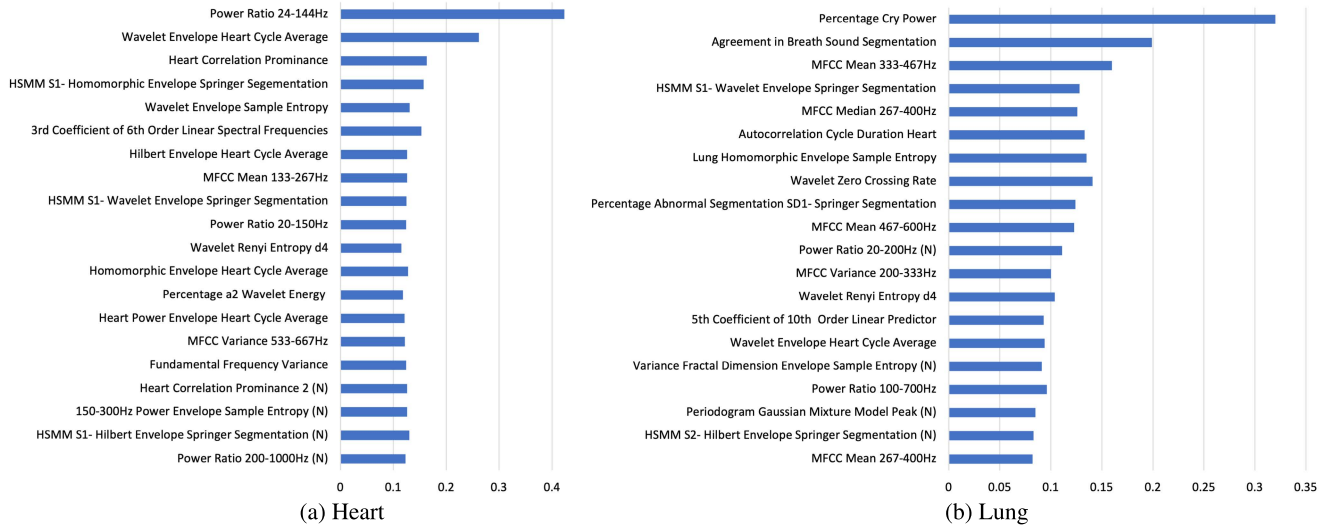
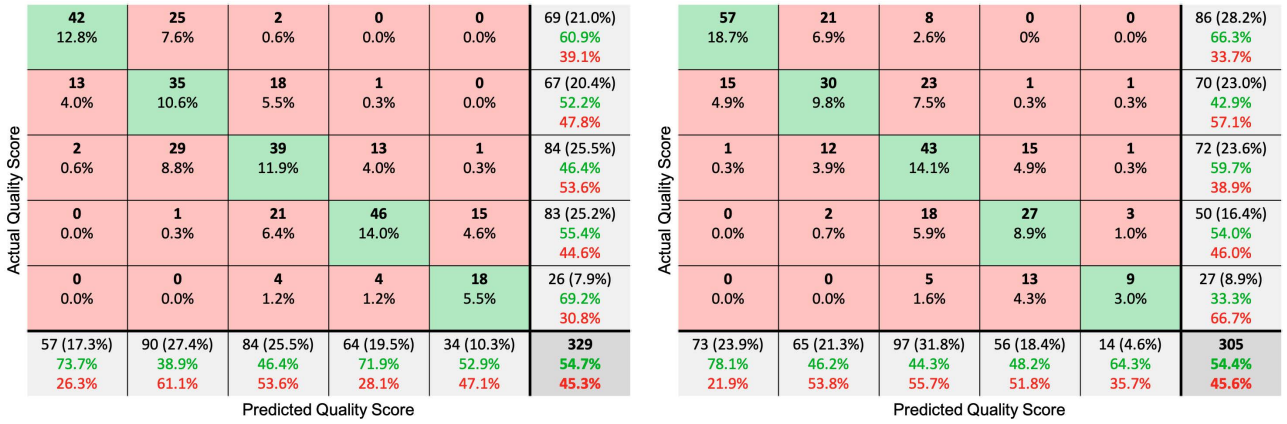


FIGURE 3. The top 20 fast features were calculated based on the mRMR algorithm with the MID method [39]. The slow features have been removed as detailed in Section IV. New features not seen in Figure 2 are labelled (N).

the classification stage, it takes 20 ms for the YAMNet-based model. As expected, YAMNet is computationally efficient due to its MobileNet architecture [44], [47].

V. RESULTS

Figures 2 and 3 show the top 20 features with all 400 features and slow features removed, respectively. Corresponding clas-



(a) Heart (b) Lung

FIGURE 4. Confusion matrix of signal quality estimation. The x-axis represents the predicted signal quality score and the y-axis represents the actual signal quality score for each chest sound recording. Green and red squares show the number of recordings that have been correctly or incorrectly predicted for that signal quality score. Additionally, the total percentage of recordings in each square is shown. Grey squares at the end of each row show sensitivity results ($\frac{\text{True Positive}}{\text{True Positive + False Negative}}$) for each actual signal quality score. Grey squares at the end of each row show precision results ($\frac{\text{True Positive}}{\text{True Positive + False Positive}}$) for each predicted signal quality score. Dark grey square summarises the information with total recordings and overall accuracy.

TABLE 3. Summary of classifier results. Heart and Lung Classifiers are trained with all features (Section II-C), whereas Heart Fast and Lung Fast Classifiers have slow features removed (Section IV). Heart and Lung YAMNet Classifiers are trained with the YAMNet deep learning model (Section II-D).

Classifier	Test MSE*	Train MSE	Test Acc (%)	Train Acc (%)	Test BAcc (%)	Train BAcc (%)
Heart	0.487	0.247	54.7	74.6	56.8	75.6
Heart Fast	0.459	0.272	55.0	72.5	56.7	74.0
Heart YAMNet	0.393	0.036	59.3	96.9	51.7	96.9
Lung	0.612	0.207	54.4	81.5	51.2	81.7
Lung Fast	0.673	0.219	47.9	81.1	46.3	81.1
Lung YAMNet	0.608	0.026	46.6	99.1	44.6	99.1

*MSE: Mean Squared Error, Acc: Accuracy, BAcc: Balanced Accuracy.

sifier results based on these top features are shown in Table 3. Heart with and without slow features perform comparably with balanced accuracy across the 5 classes being 56.8% and 56.7%, respectively. On the other hand, the removal of slow lung features results in a noticeable decrease in performance in all categories (mean squared error, accuracy and balanced accuracy). Using the deep learning YAMNet model, heart and lung signal quality classification balanced accuracy was inferior compared to both feature-based models. Heart YAMNet accuracy and mean squared error were notably better compared to feature-based models. This result can be attributed to the model tending to predict signal qualities of 2 to 4, which are more predominant in the test set.

Patient-wise cross-validation results using top 5-15 heart and 5-20 lung features are shown in Figures 4 and 6. The distinct separation of classes can be observed in the Violin plots, however, there is a large overlap between classes due to the 25-75th percentile generally varying +/−0.5 from the median. This overlap between classes is further supported in the confusion matrix results, with estimated signal quality concentrated +/−1 from the actual class and the observed accuracy of 54.7% and 54.4% for heart and lung, respectively. Top features utilised vs mean square error are shown in

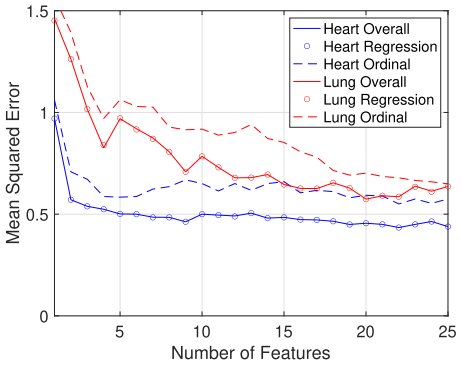
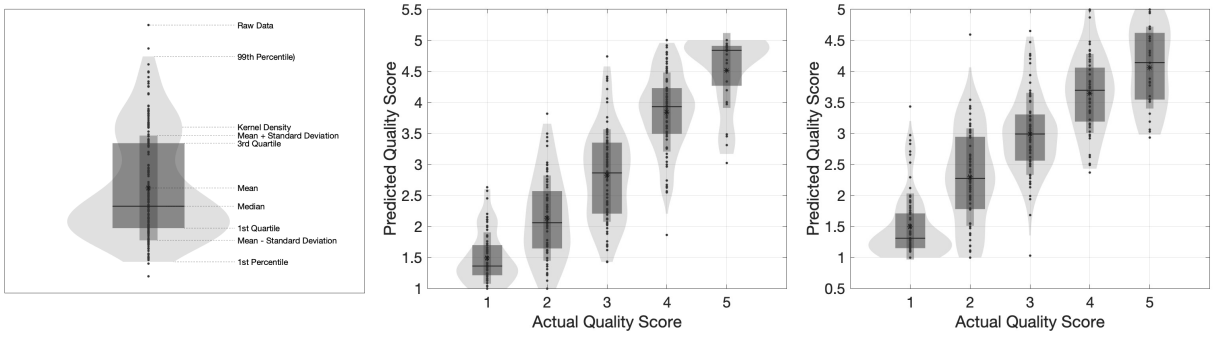


FIGURE 5. Top features utilised vs mean squared error. The top features are based on the mRMR algorithm with the MID method [39]. Classifiers are grouped as either regression based (circle) or ordinal regression based (dashed line) with results shown for heart (blue) and lung (red). Solid lines show the best performing classifier result for each feature value.

Figure 2. The best performing standard regression models outperform ordinal regression models at all number of features based on both mean squared error and accuracy.

Figure 5 shows regression classifiers outperform ordinal regression for both heart and lung signal quality classification. For patient-wise cross-validation, SVM



(a) Legend (b) Heart (c) Lung

FIGURE 6. Violin plot of signal quality estimation.

(65.5%) and KNN (34.5%) regression classifiers were the most commonly chosen for heart sound classification. Similarly, for lung sound classification, KNN (54.4%) and SVM (36.0%), Elastic-Net (6.4%) and ridge (2.4%) regression classifiers and logistic model with all thresholds (0.8%) for ordinal regression were most commonly chosen.

The regression quality assessment model used for heart and breathing rate estimation had a mean squared error of 0.505 and 0.742, and accuracy of 56.3% and 43.8% for heart and lung signal quality estimation.

Table 2 shows the mean absolute error and percentage of recordings with error less than 5 bpm for signal qualities 1 to 5. As can be seen in all cases, improvement in signal quality leads to a reduction in heart and breathing rate error, and an increase in the percentage of recordings with less than 5 bpm error. For clinical use, a mean absolute error of less than 5 bpm is typically required [52], [53]. Based on this requirement, only high-quality recordings with signal quality 5 for heart recordings and signal qualities 4 and 5 for lung recordings meet this requirement. Whereas, low-quality recordings are not appropriate for accurate vital sign estimation.

VI. DISCUSSION

In terms of modelling signal quality data, three options were available: multi-class classification, ordinal regression, and regression. As ordinal regression is a multi-class classification model that treats the classes as an ordered set that is consistent with signal quality labels, it was chosen over multi-class classification. However, knowing whether ordinal regression or standard regression is more appropriate is a more difficult task. The rationale for standard regression is signal quality makes sense as a continuous scale from 1-5 as noise volume and contamination can vary continuously. Furthermore, standard regression aids in addressing annotator disagreement shown in Figure 1. Consider 2 recordings, both with a median signal quality score of 5, but for one all 5 annotators scored the recording 5, whereas only 3 did for the other. This annotator disagreement suggests the former recording is of higher quality even though both are represented by the same score. While ordinal regression could only model in discrete classes, the benefit of standard regression of these two recordings can be scored differently more appropriately representing the actual signal quality.

However, two issues arise from using standard regression. Firstly, a signal quality estimation can go outside the range 1-5. This issue is partially addressed with signal quality being restricted to 1-5 after classification, however, this does not change the inherent method used for training the classifier itself. Secondly, whilst signal quality makes sense to be represented as a continuous value, this does not mean the discrete classes used for annotating are equally spaced. For instance, lung signal quality classes 4 and 5 sound closer to each other than signal quality classes 1 and 2 as demonstrated in Figure 6. This makes sense, as no or next to no lung sound (class 1) vs hearing partially lung sound (class 2) is an easier task than to differentiate easy to hear lung sound that both have minimal noise, typically in the form of heart sound (class 4 and 5). As shown in Figure 6, this issue is addressed, but it means that signal quality between 1-5 is not completely evenly distributed.

Based on Figure 5 standard regression outperformed ordinal regression based on mean squared error. This suggests that standard regression more appropriately represented signal quality, which fits the earlier discussion. Other contributing factors to superior performance are that continuous-valued estimation is easier to minimize mean squared error, and a larger set of regression models in comparison to ordinal regression were available. With regards to a larger set of regression models, in both python and MATLAB, regression libraries are more established, optimised and available, whereas fewer ordinal regression models are available. There is a simple method of converting several multi-class classification classifiers into an ordinal classifier, however, this is not ideal as training multiple classifiers independently is an inefficient process, and the potential for specialized algorithms that can train with a single classifier may produce superior results [42].

As shown in Table 2, high-quality (signal quality of 4 or 5) can enable accurate vital sign estimation of heart and

breathing rate for clinical usage. Whereas, low-quality recordings can provide inaccurate vital sign estimation, hindering clinical diagnosis. It can also be seen that mean absolute error increases using the Springer *et al.* heart rate method in comparison to the Schmidt *et al.* method. As the Schmidt *et al.* method is used as an initial vital sign estimation for the Springer *et al.* method's heart segmentation, the increased error suggests that poor heart rate estimation amplifies the error in the more detailed analysis of heart segmentation. Overall, improvement in signal quality can enable more accurate vital sign estimation which is necessary for clinical use and more detailed analysis.

For real-time processing, slow features namely sample entropy to autocorrelation signal, mean rate average energy and features based on Schmidt *et al.* heart segmentation, STFT envelope, singular value decomposition features were removed. The removal of features meant feature extraction times were markedly reduced from 1.12 s to 130 ms, and 2.16 s to 120 ms for heart and lung respectively. Real-time processing is less than 400 ms processing time, which is satisfied with both feature-based classifiers and the deep learning model; however, these processing times were achieved with a MacBook Pro [10]. Similar results would be expected if a desktop computer in a hospital setting or phone connected to cloud computing, whereas using phone onboard processing would be slower. Future research in investigating processing time on phones would be required to determine appropriateness. Reducing the processing times of heart and lung feature-based classifiers of 130 ms and 120 ms even further is possible. The most promising methods for reducing processing time is converting MATLAB code into optimised C code in MEX function and the other is vectorising for loops. Whereas for the deep learning model, the calculation of the Mel spectrogram and signal quality are already optimised by using in-built MATLAB functions, so minimal improvements would be expected for optimising the code. However, YAMNet did require the upsampling of the recordings to 16 kHz and the calculation of the Mel spectrogram in the 125-7500 Hz frequency range. It would be more efficient for the pre-trained YAMNet model to work for the 4 kHz sampling frequency and the calculation of the Mel spectrogram at 20-2000 Hz frequency range.

For heart sound quality classification, the removal of slow features resulted in only minor changes in results (Table 3). As only a maximum of 15 features was used, only the autocorrelation sample entropy feature was removed, which had a comparable feature selection score to other features in the top 20, meaning the removal of that feature was minor. Furthermore, the removed features important for heart classification had analogous faster features, namely, downsampled sample entropy instead of autocorrelation sample entropy, Springer *et al.* method instead of Schmidt *et al.* heart segmentation and numerous other envelope representations instead of STFT envelope. Finally, as heart results improvement plateaued after 5 features (Figure 5) the removal of features ranked 7, 17, 19, 20 would be expected to be minor.

It is noted that in Table 3 heart results with removed features perform slightly better with regards to accuracy and mean squared error, which may appear counter-initiative. Firstly, these differences are minor and secondly, as the classifiers were trained with balanced classes, hence, comparison based on balanced classes would be more appropriate. When comparing heart results with removed features with regards to balanced accuracy, it performed slightly worse as expected.

For breath sound quality classification, the removal of slow features produced a marked decrease in performance as shown in Table 3. This can be explained by the combination of higher-ranked features 2, 12, 17 being removed and up to a maximum of 20 features being used for the classifier as opposed to the heart classifier where lower-ranked features were removed and only the top 15 were features used. Additionally, there are not any features that closely resemble the mean rate average energy features that were removed. Future works in optimising mean rate average energy for real-time processing can potentially address this decrease in performance.

Heart and lung quality classifier performance achieved an accuracy of 54.7% and 54.4% and mean squared error of 0.487 and 0.612, respectively. One reason for the relatively low accuracy can be attributed to the annotator disagreement with median annotated quality differing from typically by $+/-1$ by some annotators. This annotator disagreement can be seen in Figure 1, where accuracy was 71.2% and 71.9% for heart and lung, respectively. Whilst removal of poor agreement recordings was done which improved annotator accuracy, this still suggests there is difficulty in accurately defining signal quality. In particular, annotator disagreement is high for the middle classes 2-4, which is also observed in the classifier results in Figure 6. This suggests that while annotators can generally agree on what is clearly low and high-quality, middle values are a lot harder to determine.

Potential solutions to address annotator disagreement to improve classification accuracy are the generation of an artificial dataset of heart and lung sounds with varying levels of noise. More precisely, clean heart, lung, and a variety of sources of noise such as stethoscope movement, alarms, crying and background talking can be fused together in different combinations. The signal quality label for these artificial recordings would then be the signal to noise ratio. The key benefit of the artificial dataset is a clear definition of signal quality and the generation of a balanced large number of examples of varying signal quality for the training of classifiers. However, a key question of this method is how closely these artificial recordings resemble real low and high-quality heart and lung sounds, and is there a strong correlation between signal to noise ratio and perceived signal quality by clinicians. An additional issue for the construction of an artificial dataset is obtaining enough clean heart and lung sounds, in particular lung sounds, as a majority are contaminated with heart noise. Regardless, there has been a large amount of research in artificial heart/lung recording datasets mixed either instantaneously or via convolution [13], [54]–[58].

Another contributing factor to relatively low accuracy for estimating signal quality is the small imbalanced training set. In particular, classification accuracy was only 33.3% for class 5 in the lung classifier, which contained only 8.9% of the training data. As discussed previously, given that signal quality classes 4 and 5 in the lung appear to be difficult to differentiate (Figure 6) and are underrepresented in the dataset, a larger number of recordings for training would be of benefit. A larger number of recordings would also enable a larger feature set to be utilized before over-fitting becomes a major concern. Imbalanced classes were partially addressed with data up-sampling with replacement, however, this only replicates existing recordings which can lead to an over-fitting problem. Synthetic Minority Oversampling Technique (SMOTE) may address this by generating new samples from the underrepresented class based on interpolation from the existing recordings [59]. As this is typically achieved with k=5 for k-nearest neighbours in the underrepresented class in the features space, this is not viable with the current dataset. As 5-fold cross-validation is performed, there are some folds with fewer than 5 recordings. Furthermore, with only a small number of recordings, very few new interpolated recordings could be generated.

Future collection of high-quality heart and lung sounds may address the class imbalance and improve classifier performance. However, obtaining such recordings is difficult in a noisy neonatal intensive care unit environment. One option to address this issue is the usage of more advanced denoising and sound separation techniques as opposed to standard frequency filtering. These methods can enable high-quality heart and lung sounds to be generated from noisy chest sound recordings. Non-negative matrix co-factorisation is one such method developed in previous work [15].

Similar to past work, heart classifier performance was superior to lung classifier performance as shown in Table 3 [6]. Annotator agreement both before and after removal of recordings was consistent for heart and lung, suggesting a classification of signal quality is of similar difficulty. As the majority of the features are either heart-based or have been adapted from heart features, this resulted in more suitable features for the classification of heart sound quality. Currently, there are fewer works in lung sound quality estimation, but as shown in Figures 2 and 3, tailored features of agreement in breath sound segmentation, breath sound envelope and mean rate average energy are important for lung signal quality estimation. Therefore, future work in creating further lung-based features may improve results.

With top features for signal quality classification, features with frequency ranges of 20-267 Hz and 200-467 Hz were observed for heart and lung sounds quality classifiers, respectively. This makes sense as their frequency ranges correspond closely with frequency ranges for heart and lung sounds [6]. Additionally, many heart segmentation-based features such as HSMM quality and percentage abnormal segmentation features were important for lung sound classification. As heart sounds are normally present in all lung recordings and act as noise, reducing signal quality, these heart segmentation-based features aid in determining the amount of heart noise contamination.

Results for the deep learning model for heart and lung classification were promising. Heart and lung signal quality classification balanced accuracy were 5.1% and 6.6% lower than the best feature-based models as seen in Table 3. Overall, changing the required input in two ways for YAMNet, would enable improved results. Firstly, Mel spectrogram should be calculated at 4 kHz sampling frequency and the 20-2000 Hz range. Heart sounds are still prominent in the 20-125 Hz range, hence the exclusion of this frequency band in the current YAMNet model leads to inferior results for both heart and lung signal quality classification. This is further supported in Figures 2 and 3, where various power ratios from 20-200 Hz are highly ranked features. Secondly, the input length of 0.98 s should be modified to 10 s. Whilst 0.98 s can typically obtain a couple of heartbeats consistently, this is not the case for breath sounds. With a short window length, some segments contain no breath sounds, or only partial inspiration or expiration. This means segment-based classification can vary a lot from the overall 10 s recording signal quality. In particular, faster breathing would artificially inflate the signal quality estimation for lung sounds. Additionally, the requirement of averaging out the signal quality estimation based on the 19 segments of the 10 s long recording is problematic. As signal quality is estimated in the range 1 to 5, the averaging of the 19 segments results in the tendency to predict the central signal quality values 2 to 4, as opposed to the end values of 1 and 5. To change the YAMNet model to work for a sampling frequency of 4 kHz, frequency range 20-2000 Hz and segment 10 s requires the complete retraining of the model on sound classification before transfer learning.

Another issue with the deep learning model was overfitting, as can be seen with the high train set balanced accuracy of 96.9% and 99.1% for heart and lung sound quality classification. This issue was attempted to be addressed in two ways. Firstly, the maximum epochs were restricted to 30, with the best performance occurring with 10 epochs. Additionally, instead of all layers in the deep learning model being updated, the first 40 and 60 layers of the model had their weights frozen for heart and lung signal quality classification. In future, patient-wise hyperparameter optimisation on a larger set of hyperparameters values for max epochs and number of layers with frozen weights, and additional hyperparameters such as learning rate and learning rate drop-off, may enable improved results. However, similar to the complete retraining of the YAMNet model before transfer learning, these are highly computationally expensive processes. Additionally, a larger dataset would aid in preventing overfitting.

VII. CONCLUSION

Stethoscope-recorded chest sounds provide affluent information about neonatal health status, in particular for cardio-respiratory health assessment. In combination with telehealth, digital stethoscopes can increase the availability

of quality healthcare for the early diagnosis and prognosis of newborns. However, as shown in this paper, the acquisition of high-quality recordings is necessary to obtain accurate vital signs for clinical use. In order to achieve this, accurate signal quality assessment is required for both heart and lung sounds recorded from the digital stethoscope. Signal quality assessment enables feedback to non-expert users on the quality of recordings and aids the clinical decision support system for automated analysis of those recordings. This paper presented a newborn-focused automatic heart and lung sound quality assessment on a five-level quality scale using a variety of regression methods and a deep learning YAMNet model. Overall, for the best-performing classifiers, heart and lung quality were estimated with a mean squared error of 0.487 and 0.612, taking 1.12 s and 2.16 s to compute per recording, respectively. For real-time application, heart and lung quality were estimated in under 130 ms with the mean square error of 0.459 and 0.673, respectively. The removal of the need for hand-crafted features and utilising deep learning showed promising results, with future work in improved pre-training for transfer learning required.

ACKNOWLEDGMENT

The author 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] *Neonatal Mortality—UNICEF DATA*. Accessed: Oct. 5, 2019. [Online]. Available: https://data.unicef.org/topic/child-survival/neonatal-mortality/
[2] United Nations Department of Economic and Social Affairs. *Goal 3: Ensure Healthy Lives and Promote Well-Being for All at All Ages*. Accessed: Jan. 3, 2021. [Online]. Available: https://sdgs.un.org/goals/goal3
[3] A. Ramanathan, L. Zhou, F. Marzbanrad, R. Roseby, K. Tan, A. Kevat, and A. Malhotra, "Digital stethoscopes in paediatric medicine," *Acta Paediatrica*, vol. 108, no. 5, pp. 814–822, May 2019.
[4] A. C. Kevat, D. V. R. Bullen, P. G. Davis, and C. O. F. Kamlin, "A systematic review of novel technology for monitoring infant and newborn heart rate," *Acta Paediatrica*, vol. 106, no. 5, pp. 710–720, May 2017.
[5] 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, Apr. 2020.
[6] E. Grooby, J. He, J. Kiewsky, D. Fattahi, L. Zhou, A. King, A. Ramanathan, A. Malhotra, G. A. Dumont, and F. Marzbanrad, "Neonatal heart and lung sound quality assessment for robust heart and breathing rate estimation for telehealth applications," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 12, pp. 4255–4266, Dec. 2021.
[7] 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, Jan. 2015.
[8] D. B. Springer, T. Brennan, N. Ntusi, H. Y. Abdelrahman, L. J. Zühlke, B. M. Mayosi, L. Tarassenko, and G. D. Clifford, "Automated signal quality assessment of mobile phone-recorded heart sound signals," *J. Med. Eng. Technol.*, vol. 40, nos. 7–8, pp. 342–355, Nov. 2016.
[9] K. Shi, S. Schellenberger, F. Michler, T. Steigleder, A. Malessa, F. Lurz, C. Ostgathe, R. Weigel, and A. Koelpin, "Automatic signal quality index determination of radar-recorded heart sound signals using ensemble classification," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 3, pp. 773–785, Mar. 2020.
[10] C. E. Valderrama, F. Marzbanrad, L. Stroux, B. Martinez, R. Hall-Clifford, C. Liu, N. Katebi, P. Rohloff, and G. D. Clifford, "Improving the quality of point of care diagnostics with real-time machine learning in low literacy LMIC settings," in *Proc. 1st ACM SIGCAS Conf. Comput. Sustain. Soc.*, Jun. 2018, pp. 1–11.
[11] I. Grzegorczyk, M. Soliński, M. Łepek, A. Perka, J. Rosiński, J. Rymko, K. Stępień, and J. Gierałtowski, "PCG classification using a neural network approach," in *Proc. Comput. Cardiol. Conf. (CinC)*, Sep. 2016, pp. 1129–1132.
[12] Q.-U.-A. Mubarak, M. U. Akram, A. Shaukat, 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, Oct. 2018.
[13] A. Kala, A. Husain, E. D. McCollum, and M. Elhilali, "An objective measure of signal quality for pediatric lung auscultations," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2020, pp. 772–775.
[14] D. Emmanouilidou, E. D. Mccollum, D. E. Park, and M. Elhilali, "Computerized lung sound screening for pediatric auscultation in noisy field environments," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 7, pp. 1564–1574, Jul. 2018.
[15] E. Grooby, J. He, D. Fattahi, L. Zhou, A. King, A. Ramanathan, A. Malhotra, G. A. Dumont, and F. Marzbanrad, "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. (EMBC)*, Nov. 2021, pp. 5668–5673.
[16] D. Chakraborty, S. Bhattacharya, A. Thakur, A. R. Gosthipaty, and C. Datta, "Feature extraction and classification of phonocardiograms using convolutional neural networks," in *Proc. IEEE 1st Int. Conf. Converg. Eng. (ICCE)*, Sep. 2020, pp. 275–279.
[17] L. Zhou, F. Marzbanrad, A. Ramanathan, D. Fattahi, P. Pharande, and A. Malhotra, "Acoustic analysis of neonatal breath sounds using digital stethoscope technology," *Pediatric Pulmonol.*, vol. 55, no. 3, pp. 624–630, Mar. 2020.
[18] A. Ramanathan, F. Marzbanrad, K. Tan, F.-T. Zohra, M. Acchiardi, R. Roseby, A. Kevat, and A. Malhotra, "Assessment of breath sounds at birth using digital stethoscope technology," *Eur. J. Pediatrics*, vol. 179, pp. 781–789, Jan. 2020.
[19] J. L. Fleiss, "Measuring nominal scale agreement among many raters," *Psychol. Bull.*, vol. 76, no. 5, p. 378, 1971.
[20] M. Abdollahpur, A. Ghaffari, S. Ghiasi, and M. J. Mollakazemi, "Detection of pathological heart sounds," *Physiol. Meas.*, vol. 38, no. 8, p. 1616, 2017.
[21] H. Naseri and M. R. Homaeinezhad, "Computerized quality assessment of phonocardiogram signal measurement-acquisition parameters," *J. Med. Eng. Technol.*, vol. 36, no. 6, pp. 308–318, Aug. 2012.
[22] M. Zabihi, A. B. Rad, S. Kiranyaz, M. Gabbouj, and A. K. Katsaggelos, "Heart sound anomaly and quality detection using ensemble of neural networks without segmentation," in *Proc. Comput. Cardiol. Conf. (CinC)*, Sep. 2016, pp. 613–616.
[23] 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, Feb. 2021, Art. no. 7565398.
[24] T. Li, H. Tang, T. Qiu, and Y. Park, "Best subsequence selection of heart sound recording based on degree of sound periodicity," *Electron. Lett.*, vol. 47, no. 15, pp. 841–843, 2011.
[25] D. Kumar, P. Carvalho, M. Antunes, R. Paiva, and J. Henriques, "Noise detection during heart sound recording using periodicity signatures," *Physiol. Meas.*, vol. 32, no. 5, p. 599, 2011.
[26] J. Gierałtowski, K. Ciuchciński, I. Grzegorczyk, K. Kośna, M. Soliński, and P. Podziemski, "RS slope detection algorithm for extraction of heart rate from noisy, multimodal recordings," *Physiol. Meas.*, vol. 36, no. 8, p. 1743, 2015.
[27] 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.
[28] 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, p. 513, 2010.
[29] F. Beritelli and A. Spadaccini, "Heart sounds quality analysis for automatic cardiac biometry applications," in *Proc. 1st IEEE Int. Workshop Inf. Forensics Secur. (WIFS)*, Dec. 2009, pp. 61–65.
[30] C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 5, no. 1, pp. 3–55, 2001.
[31] H. Liang, S. Lukkarinen, and I. Hartimo, "Heart sound segmentation algorithm based on heart sound envelogram," in *Proc. Comput. Cardiol.*, 1997, pp. 105–108.
[32] Y. L. Yap and Z. Moussavi, "Respiratory onset detection using variance fractal dimension," in *Proc. 23rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, vol. 2, Oct. 2001, pp. 1554–1556.

[33] S. Huq and Z. Moussavi, "Automatic breath phase detection using only tracheal sounds," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol.*, Aug. 2010, pp. 272–275.
[34] S. Huq and Z. Moussavi, "Acoustic breath-phase detection using tracheal breath sounds," *Med. Biol. Eng. Comput.*, vol. 50, no. 3, pp. 297–308, Mar. 2012.
[35] Z. K. Moussavi, M. T. Leopando, H. Pasterkamp, and G. Rempel, "Computerised acoustical respiratory phase detection without airflow measurement," *Med. Biol. Eng. Comput.*, vol. 38, no. 2, pp. 198–203, Mar. 2000.
[36] The Royal Children's Hospital Melbourne. (Jul. 2020). *Clinical Practice Guidelines: Acceptable Ranges for Physiological Variables*. Accessed: Jan. 20, 2021. [Online]. Available: https://www.rch.org.au/clinicalguide/guideline_index/Normal_Ranges_for_Physiological_Variables/
[37] S. Fleming, M. Thompson, R. Stevens, C. Heneghan, A. Plüddemann, I. Maconochie, L. Tarassenko, and D. Mant, "Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: A systematic review of observational studies," *Lancet*, vol. 377, no. 9770, pp. 1011–1018, 2011.
[38] *GitHub—Heart and Lung Signal Quality Estimation*. Accessed: Sep. 9, 2021. [Online]. Available: https://github.com/egrooby-monash/Heart-and-Lung-Signal-Quality-Estimation/
[39] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.
[40] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Nov. 2011.
[41] F. Pedregosa-Izquierdo, "Feature extraction and supervised learning on fMRI: From practice to theory," Ph.D. dissertation, Doctoral School Comput. Sci., Univ. Pierre et Marie Curie-Paris VI, Paris, France, 2015.
[42] E. Frank and M. Hall, "A simple approach to ordinal classification," in *Proc. Eur. Conf. Mach. Learn.* Berlin, Germany: Springer, 2001, pp. 145–156.
[43] C. Sitaula, J. He, A. Priyadarshi, M. Tracy, O. Kavehei, M. Hinder, A. Withana, A. McEwan, and F. Marzbanrad, "Neonatal bowel sound detection using convolutional neural network and Laplace hidden semi-Markov model," 2021, *arXiv:2108.07467*.
[44] S. Hershey, S. Chaudhuri, D. P. W. Ellis, J. F. Gemmeke, A. Jansen, R. C. Moore, M. Plakal, D. Platt, R. A. Saurous, B. Seybold, M. Slaney, R. J. Weiss, and K. Wilson, "CNN architectures for large-scale audio classification," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 131–135.
[45] J. F. Gemmeke, D. P. W. Ellis, D. Freedman, A. Jansen, W. Lawrence, R. C. Moore, M. Plakal, and M. Ritter, "Audio set: An ontology and human-labeled dataset for audio events," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 776–780.
[46] Sound and Video Teams. *Audioset*. Accessed: Oct. 18, 2021. [Online]. Available: https://research.google.com/audioset/
[47] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.
[48] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[49] *Infinity M540*. Accessed: Dec. 9, 2021. [Online]. Available: https://www.draeger.com/en_aunz/Products/Infinity-M540-monitor
[50] *Fast, Slow and Irregular Heartbeats (Arrythmia)*. Accessed: Jul. 31, 2020. [Online]. Available: https://www.healthychildren.org/English/health-issues/conditions/heart/Pages/Irregular-Heartbeat-Arrhythmia.aspx
[51] *Normal Heart Rates for Children—Children's Health*. Accessed: Jul. 31, 2020. [Online]. Available: https://www.childrens.com/health-wellness/is-your-childs-heart-rate-healthy
[52] D. B. Springer, T. Brennan, J. Hitzeroth, B. M. Mayosi, L. Tarassenko, and G. D. Clifford, "Robust heart rate estimation from noisy phonocardiograms," in *Proc. Comput. Cardiol.*, 2014, pp. 613–616.
[53] S. Nizami, A. Bekele, M. Hozayen, K. J. Greenwood, J. Harrold, and J. R. Green, "Measuring uncertainty during respiratory rate estimation using pressure-sensitive mats," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 7, pp. 1535–1542, Jul. 2018.
[54] G. Shah, P. Koch, and C. B. Papadias, "On the blind recovery of cardiac and respiratory sounds," *IEEE J. Biomed. Health Informat.*, vol. 19, no. 1, pp. 151–157, Jan. 2015.
[55] C. Lin and E. Hasting, "Blind source separation of heart and lung sounds based on nonnegative matrix factorization," in *Proc. Int. Symp. Intell. Signal Process. Commun. Syst.*, Nov. 2013, pp. 731–736.
[56] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," *IEEE Trans. Audio, Speech, Language Process.*, vol. 14, no. 4, pp. 1462–1469, Jul. 2006.
[57] A. G. Rudnitskii, "Using nonlocal means to separate cardiac and respiration sounds," *Acoust. Phys.*, vol. 60, no. 6, pp. 719–726, Nov. 2014.
[58] 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.
[59] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002.

E. GROOBY (Member, IEEE) received the Bachelor of Biomedicine and the Master of Engineering (biomedical) degrees from The University of Melbourne, Melbourne, VIC, Australia, in 2017 and 2019, respectively. He is currently pursuing the joint Ph.D. degree in electrical and computer systems engineering with Monash University, Melbourne, and The University of British Columbia, Vancouver, BC, Canada. From 2017 to 2019, he was a Research Student at the Walter and Eliza Hall Institute of Medical Research, Peter MacCallum Cancer Centre and Defence Science and Technology Group. From 2019 to 2020, he was a Research Engineer at Cochlear. His research interests include biomedical signal processing and medical device development.

C. SITAULA received the Ph.D. degree from Deakin University, Geelong, VIC, Australia, in 2021. He worked in industry and academia in Nepal for a number of years before joining Deakin University for his Ph.D. degree. He is currently a Research Fellow with the Department of Electrical and Computer Systems Engineering, Monash University, Melbourne, VIC, Australia. He has published research articles in top-tier conferences and journals in the field of deep learning and machine learning. His research interests include computer vision, signal processing, and machine learning.

D. FATTAHI is currently pursuing the Ph.D. degree in biomedical engineering (bioelectric) with Shiraz University, Shiraz, Iran. From March 2019 to September 2019, he spent a Research Internship at the Biomedical Signal Processing Laboratory, Department of Electrical and Computer Systems Engineering, Monash University, Australia. His research interests include biomedical signal processing (ECG, EEG, PCG, and lung sound), parameter estimation, time-frequency analysis, and blind source separation.

R. SAMENI (Senior Member, IEEE) received the bachelor's degree in electronics engineering from Shiraz University, Iran, in 2000, the master's degree in biomedical engineering from the Sharif University of Technology, Iran, in 2003, and the dual Ph.D. degrees in signal processing and biomedical engineering from the Institut National Polytechnique de Grenoble (INPG), France, and the Sharif University of Technology, in 2008. He was a Tenured Associate Professor with the School of Electrical and Computer Engineering, Shiraz University (2008–2018), and an Invited Senior Researcher at the GIPSA-laboratory, Grenoble, France (2018–2020). Since 2020, he has been an Associate Professor of biomedical engineering at Emory University, Atlanta, GA, USA. His research interests include statistical signal processing with a special interest in mathematical modeling and analysis of biomedical systems and signals.

K. TAN is currently a Consultant Neonatologist at Monash Children's Hospital and an Adjunct Associate Professor at Monash University. His research interests include clinical registry, clinical trials, big data analysis, and clinical practice improvement.

L. ZHOU received the M.B.B.S. degree, in 2012. He is currently pursuing the Ph.D. degree investigating umbilical cord blood-derived cell therapies for preterm brain injury with Monash University. He is also a Neonatologist at Monash Children's Hospital, Melbourne, VIC, Australia, completing his Neonatal Fellowship Training, in 2020.

A. KING received the Bachelor of Medical Science (Hons.) and the Bachelor of Medical Science and Doctor of Medicine degrees from Monash University, Melbourne, VIC, Australia, in 2019 and 2020, respectively. She is a current paediatric trainee at the Royal Children's Hospital, Melbourne and holds an affiliate position with the Department of Paediatrics, Monash University.

A. RAMANATHAN received the Bachelor of Medical Science and M.B.B.S. degrees (Hons.) from Monash University, Melbourne, VIC, Australia. He is currently a Resident Medical Officer at Perth Children's Hospital.

A. MALHOTRA received the M.D. and Ph.D. degrees. He is currently a Senior Neonatologist at Monash Children's Hospital, and an Associate Professor (Research)/NHMRC Fellow at Monash University, Melbourne, VIC, Australia. He has a large research program, with interests in neonatal lung and brain injury, with more than $7 million in research funding. He has published more than 100 peer-reviewed articles, and four book chapters to date. Together with Dr. Fae Marzbanrad, their team researches digital health technologies to improve neonatal cardiorespiratory monitoring.

G. A. DUMONT (Life Fellow, IEEE) received the Dipl.Ing. degree from the Ecole Nationale Supérieure d'Arts et Metiers, Paris, France, in 1973, and the Ph.D. degree in electrical engineering from McGill University, Montreal, QC, Canada, in 1977. He was with Tioxide, France, from 1973 to 1974, and again from 1977 to 1979. He was with Paprican, from 1979 to 1989, first in Montreal and then in Vancouver. In 1989, he joined the Department of Electrical and Computer Engineering, The University of British Columbia, where he is currently a Professor and a Distinguished University Scholar. From 2000 to 2002, he was the Associate Dean of the Research for the Faculty of Applied Science. Since 2008, he has been an Associate Member of the UBC Department of Anesthesiology Pharmacology and Therapeutics. He is a Principal Investigator at the BC Children's Hospital Research Institute and the Co-Founder and the Co-Director of the Digital Health Innovation Laboratory (DHIL). His current research interests include patient monitoring, signal processing for physiological monitoring, and physiological closed-loop control systems, such as automated drug delivery in anesthesia, circadian rhythms, global and mobile health, non-contact patient vital sign assessment, and brain monitoring via electroencephalography and near-infrared spectrometry. He was awarded the 1979 IEEE Transactions on Automatic Control Honorable Paper Award; the 1985 Paprican Presidential Citation; the 1990 UBC Killam Research Prize; the 1995 CPPA Weldon Medal; the 1998 Universal Dynamics Prize for Leadership in Process Control Technology; the IEEE Control Systems Society 1998 Control Systems Technology Award; three NSERC Synergy Awards, the latest one in 2016 for the Development of the Phone Oximeter; the 2010 Brockhouse Canada Prize for Interdisciplinary Research in Science and Engineering. From 2011 to 12, and again from 2018 to 2019, he was a UBC Peter Wall Distinguished Scholar in Residence. In 2020, he was awarded the IEEE Control Systems Society Transition to Practice Award. In 2017, he was elected a fellow of the International Federation of Automatic Control as well as a fellow of the Royal Society of Canada.

F. MARZBANRAD (Senior Member, IEEE) received the Ph.D. degree from the University of Melbourne, Australia, in 2016. She is currently a Lecturer and the Head of the Biomedical Signal Processing Laboratory, Department of Electrical and Computer Systems Engineering, Monash University, Australia. Her research interests include biomedical signal processing, machine learning, affordable medical technologies, and mobile-health.