Direct machine learning reconstruction of respiratory variation waveforms from resting state fMRI data in a pediatric population

Abdoljalil Addeh [a,b,c,e], Fernando Vega [a,b,c,e], Prathistith Raj Medi [f], Rebecca J. Williams [c,d,e], G. Bruce Pike [c,d,e], M. Ethan MacDonald [a,b,c,e,*]

[a] Department of Biomedical Engineering, Schulich School of Engineering, University of Calgary, Canada
[b] Department of Electrical & Software Engineering, Schulich School of Engineering, University of Calgary, Canada
[c] Department of Radiology, Cumming School of Medicine, University of Calgary, Canada
[d] Department of Clinical Neurosciences, Cumming School of Medicine, University of Calgary, Canada
[e] Hotchkiss Brain Institute, Cumming School of Medicine, University of Calgary, Canada
[f] Data Science and Artificial Intelligence, International Institute of Information Technology, Naya Raipur, India

ARTICLE INFO

ABSTRACT

In many functional magnetic resonance imaging (fMRI) studies, respiratory signals are unavailable or do not have acceptable quality due to issues with subject compliance, equipment failure or signal error. In large databases, such as the Human Connectome Projects, over half of the respiratory recordings may be unusable. As a result, the direct removal of low frequency respiratory variations from the blood oxygen level-dependent (BOLD) signal time series is not possible. This study proposes a deep learning-based method for reconstruction of respiratory variation (RV) waveforms directly from BOLD fMRI data in pediatric participants (aged 5 to 21 years old), and does not require any respiratory measurement device. To do this, the Lifespan Human Connectome Project in Development (HCP-D) dataset, which includes respiratory measurements, was used to both train a convolutional neural network (CNN) and evaluate its performance. Results show that a CNN can capture informative features from the BOLD signal time course and reconstruct accurate RV timeseries, especially when the subject has a prominent respiratory event. This work advances the use of direct estimation of physiological parameters from fMRI, which will eventually lead to reduced complexity and decrease the burden on participants because they may not be required to wear a respiratory bellows.

1. Introduction

In recent years, fMRI has shown promise for studying pediatric cohorts to understand brain function both in typically developing children and those with neurodevelopmental disorders. However, the potential of fMRI is limited by difficulties associated with its use in children. It is common for children to be anxious, claustrophobic, and restless during an fMRI scan, limiting compliance. Furthermore, children may have difficulty following instructions and comprehending requirements during imaging (Rassler et al., 2022; Shechner et al., 2013). In adults, a brief or limited fMRI preparation session may be sufficient, but fMRI preparation for children often needs to be more extensive and individualised.

Hemodynamics in the brain are complex, particularly with disease (MacDonald et al., 2016). Slow variations in breathing rate and depth during fMRI scanning can alter cerebral blood flow and consequently the blood oxygen level-dependent (BOLD) signal (MacDonald et al., 2018, 2020; Williams et al., 2021). Correcting for the effects of low frequency respiratory fluctuations has been shown to improve

the detection of task-activated voxels and reduce false−positives and false−negatives in resting-state functional connectivity (Birn et al., 2006; Chang and Glover, 2009; Chu et al., 2018; Golestani and Chen, 2020; Golestani et al., 2016). In general, attempts to gauge respiratory contributions to fMRI signals require a respiratory measurement recorded during scanning. In many studies, particularly in pediatric populations, physiological signals (respiratory and cardiac) are not recorded or have poor quality, due to issues with participant compliance (*e.g.*, moving or disconnecting the measuring equipment).

Over the past decades several methods such as a notch filter (Biswal et al., 1996), k-space domain (Frank et al., 2001; Glover and Lai, 1998; Hu et al., 1995; Wowk et al., 1997), image domain (Chuang and Chen, 2001; Glover et al., 2000), and convolution model-based approaches (Birn et al., 2006; Chang and Glover, 2009; Golestani et al., 2015; Wise et al., 2004) have been developed for reducing the effect of physiological noise in fMRI time series. In notch filter methods, the fundamental frequency of the respiratory signal is identified from external recordings and used to filter those components

of the BOLD signal. Glover et al. proposed an image-based respiratory effect cancellation method named 'retrospective image correction technique' or RETROICOR (Glover et al., 2000). RETROICOR models physiological noise using a Fourier expansion based on the phases of the externally recorded respiratory or cardiac signals. Compared to a notch filter that works only on a prescribed range of frequencies sufficiently sampled in the time series, RETROICOR can account for fluctuations that are of a higher frequency. However, RETROICOR does not remove slow (~0.03 Hz or lower) signal changes induced by breath-to-breath variations in the depth and rate of breathing. Consequently, more sophisticated model-based approaches have been proposed to account for the effects of low-frequency respiration variations extracted from external respiratory measurements (Birn et al., 2008; Chang and Glover, 2009; Golestani et al., 2015; Power et al., 2020). While these techniques perform very well, they require the collection of external signals that can be cumbersome and prone to error in the MR environment, particularly in pediatric populations.

Neural network algorithms utilize large volumes of data and discover specific trends and patterns that may not be apparent to human experts. In a study by Salas and colleagues, a convolutional neural network (CNN) and single-unit dense layer linear network were used to reconstruct respiratory signal variation from fMRI data (Salas et al., 2021). In two studies conducted by Bayrak et al., a U-Net was utilized for RV reconstruction (Bayrak et al., 2020), and a deep long-short term memory network was used for joint estimation of RV and heart rate (Bayrak et al., 2021). The developed methods were tested on an adult population and showed a great deal of success. There were some specific limitations noted about their work, in particular: BOLD signals were band-pass filtered in a range of 0.01–0.15 Hz and temporally downsampled by a factor of 2, the study was limited only to the adult population, and smoothing was done on the respiratory variation signals before trying to do the estimation. The typical respiratory rate in adults is within the range of 10–20 breaths per min (0.16–0.33 Hz), but there are large number of periods in the Human Connectome Project Young Adults (HCP-YA) (dataset used by Salas and colleagues) where participants had slow breathing, in range (0.07–0.16 Hz). Temporal smoothing makes it difficult to pinpoint the precise moment of activity and limited the performance of the CNN. Moreover, due to edge-effects, these models are unable to reconstruct the entire RV signal and some information is discarded at the beginning and end of the scan. Finally, extensions to their work in pediatric populations are particularly important, as children are known to have faster and more variable respiration, more motion and lower compliance with respiratory measurement equipment.

This study proposes a novel CNN architecture to reconstruct the full respiratory signal variation waveforms from fMRI data alone in a pediatric population, with minimum preprocessing and smoothing both on the BOLD signals and respiratory belt traces. In our work, we show that many of the respiratory signals available in the Human Connectome Project – Developmental (HCP-D) database are unsatisfactory for training because of quality, further highlighting the need for such an algorithm. We trained the model using a 10-fold cross validation procedure with performance metrics. Correlation as a performance metric is shown to have limitations and so several performance metrics as used. Finally, we do an ablation study (i.e., test performance against model parameters) to show the model performance as a function of the CNN kernel size and position. The advancements on the previous works by Salas et al. (2021) and Bayrak et al. (2021, 2020), include the more sophisticated CNN model, less pre-processing of the fMRI data, application to pediatric population where the resting state scan is half the length, and using more extensive evaluation criterion.

2. Materials and methods

2.1. Dataset

The HCP-D dataset is a publicly available and has been comprehensively described (Harms et al., 2018; Somerville et al., 2018). We give a brief overview here. The HCP-D aims to enroll 1300 healthy children, adolescents, and young adults ranging in age range between 5 and 21 years from diverse geographical, ethnic, and socioeconomic backgrounds. Brain imaging is performed on a 3T Siemens Prisma platform and includes structural, functional (resting state and task-based), diffusion, and perfusion imaging, and physiological signal monitoring. Data is acquired during two separate MRI sessions.

In this paper, we focus on the resting state fMRI scans and the associated respiration measurements. For the resting state fMRI scans, subjects were scanned at rest 4 times, each time for 6.41 min (two scans per day on two days). Sequences were counterbalanced with anterior to posterior (AP)/posterior to anterior (PA) phase encoding, with TR=800 ms, TE=37 ms, flip angle=52°, with a 32-channel head coil, multi-band factor 8, and 2 mm isotropic voxels with full brain coverage. During HCP-D resting-state fMRI scanning, participants are instructed to remain still, stay awake, and blink normally while looking at a fixation crosshair. Table 1 shows the details of HCP-D resting-state fMRI data. Compared to HCP-YA project where the scanning duration of resting-state fMRI was 15 min, in HCP-D project it is shortened to ~7 min to improve tolerability for younger participants.

The HCP-D dataset has the unique advantage of containing four prolonged resting state scans that allows us to examine the changes in respiratory measurements occurring over the course of a scanning session. Respiratory data was acquired via the Siemens Physiology Monitoring Unit (PMU). The respiratory recording is obtained via a pneumatic hose connected to a respiratory cushion placed under an elastic belt strapped around the subject's abdomen, and output is in arbitrary units with a 400 Hz sampling frequency. In this study, fMRI data from 614 subjects from HCD0001305 to HCD2996590 (2451 scans in total) were investigated, and 306 scans belonging to 215 subjects were selected based on the quality of their respiratory signals.

2.2. Quality control of respiratory signals

All respiratory belt traces were first visually inspected and processed with a custom program to identify partial or complete corruption. There are several potential sources of variation arising from the participant in the respiration recordings, such as: talking, breath holding, hyperventilation and hypoventilation. Fig. 1a shows an example of the impact of talking on normal respiratory signal. In addition to breathing variations, other technical problems such as improper attachment of the belt to the

Table 1

Resting-state fMRI imaging protocol in HCP-D. The four 6:41minute resting-state fMRI runs (26-minute total) are acquired in the two separate fMRI sessions over two days (Harms et al., 2018).

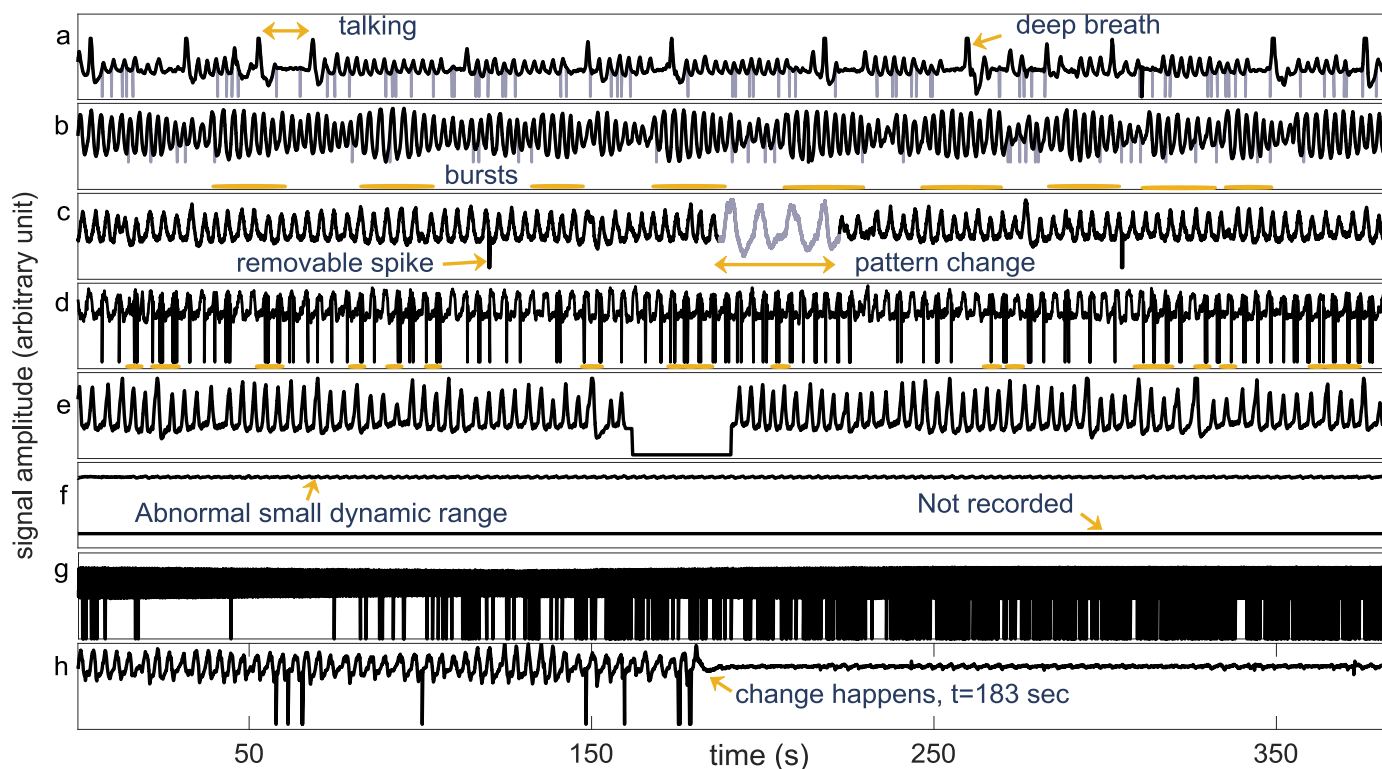| Session	Modality	Scan	Resolution (mm)	fMRI volumes	Duration (min:sec)	Participant action
	Spin echo field maps	AP & PA	2		0.18	NA
1	BOLD Resting state	Run 1 AP	2	488	6:41	Fixation
	BOLD Resting state	Run 2 PA	2	488	6:41	Fixation
	Spin echo field maps	AP & PA	2		0.18	NA
2	BOLD Resting state	Run 1 AP	2	488	6:41	Fixation
	BOLD Resting state	Run 2 PA	2	488	6:41	Fixation

Fig. 1. Example of respiratory signals from different groups. A, b, c) Signal with removable spikes (HCD0425335, session 1, run 2; HCD0271031, session 2, run 2; HCD2000111, session 2, run 1). Light grey graphs show the original respiratory signal with removable spikes, d) Signal with unremovable spikes marked by color dashes (HCD2335344, session 2, run 2), e) partially recorded signal (HCD0968878, session 2, run 2). In this example, spikes are removed from recorded parts of the signal to show the impact of spike elimination clearly, f) Not recorded (signal with zero amplitude, HCD0110411, session 2, run 1) and signal varies only in a small range having square pulse-shape pattern (HCD0694564, session 1, run 2), g) very distorted signals with high frequency noise (HCD0146937, session 1, run 1), h) connections changed at a certain point (HCD1796577, session 1, run 1).

transducer or the transducer to the subject, and improper fitting the belt around subject's abdomen, can corrupt the data completely or partially. Moving around (*i.e.*, squirming) or pulling on the respiration belt are more common in pediatric populations and represents another source of error. Depending on the severity of corruption, outlier replacement can be used to correct spurious spike artifacts. Three examples of spurious spike artifacts that can be eliminated using outlier replacement are shown by Fig. 1a–c. In Fig. 1, light grey graph shows the original respiratory signal including spikes and the black graph shows the signal after spike correction using the MATLAB 'filloutliers (signal, 'previous', 'movmedian', window_size)' command. Local outliers are defined as elements more than three scaled median absolute deviations in a moving window. The breathing rate in HCP-D dataset is 0.315 ± 056 Hz, which means breaths have ~3 s duration as expected in children. As outliers occur repeatedly and at random places in a breath, it is not reasonable to use a large window covering a whole breath duration (considering 400 Hz sampling rate, a window size 1200 is required to cover a breath). Large differences between minimum and maximum values of belt traces at the end of inhalation and exhalation can put some obvious outliers inside the accepted range (smaller than three local standard deviations away from the local median within a window) and the algorithm will miss those outliers. Therefore, the largest length of moving window set to 1 s (or 400 points as the sampling rate is 400 Hz). These investigations revealed that for each signal a certain window length will lead to the best outlier replacement (minimum number of spikes), and by decreasing or increasing the window length the outlier replacement will fail again. More details are provided in the Supplementary Material. The length of the moving window is selected for each signal separately, with the goal of obtaining a minimum number of spikes. Deep breathes in Fig. 1a, serial tapering of respiratory depth spanning 45 s at a time or bursts in

Fig. 1b, and significant change in breathing depth and rate in Fig. 1c, and any change in breathing pattern has distinct influences on BOLD signals and signal covariance. If a RV timeseries is available, the effect of the breathing pattern can be accounted for in the fMRI data analysis to reduce this confound.

All respiratory signals were evaluated and classified as acceptable or into one of 7 classes of corruption: Class 1: Signals with removable spurious spike artifacts (Fig. 1a to c). All HCP-D respiratory time courses have some amount of spurious spike artifacts. For signals belonging to Class 1, the outlier replacement algorithm was able to remove all spurious spike artifacts completely. 352 out of all 2451 respiratory signals are class 1 (14.36%). Class 2: Signals with unremovable spurious spike artifacts (Fig. 1d). The analyses showed that all the signals that have unremovable spikes (Class 2) have more than 100 spikes after outlier replacement. Spikes are a significant element in a respiratory signal because they impact the standard deviation of the signal (or RV) substantially. Even one spike in the respiratory signal can create a bump in RV waveform and mislead the model training. 1849 out of all 2451 signals are class 2 (75.43%). Class 3: Partially recorded signal (*e.g.*, Fig. 1e) occurred 130 out of 2451 times (5.3%). One possible reason for this type of error can be subject movement and displacement of the belt or disconnection of the transducer. Class 4: Signals with very low amplitude (Fig. 1f) were found in 65 out of 2451 scans (2.65%). Normally, recorded respiratory signal in the HCP-D project vary in range 668 ± 590 and 3306 ± 534 (arbitrary units). Signals in Class 4 have a dynamic range smaller than 100 resulting in a square pulse-shape pattern. A square pulse-shape pattern is generally not observed in signals with dynamic range higher than 100. More details on respiratory signals assigned as Class 4 are provided in the Supplementary Material. Class 5: Not recorded cases (Fig. 1f) constitute 15 out of 2451 signals (0.61%).

Class 6: Very distorted signals with high frequency noise (Fig. 1g) – constitute 24 out of all 2451 (0.97%). More details on respiratory signals assigned as Class 6 are provided in the Supplementary file. Class 7: Connections changed at a certain point (Fig. 1h). 16 out of 2451 signals are classified as class 7 (0.65%). Many of the recorded respiratory signals in the HCP-D dataset don't have acceptable quality and cannot be used for further analysis, and particularly not for training a machine learning model as we do not want the model to learn these erroneous signals. As respiratory signals in Class 1 have good quality, the respiratory signals in this class and their associated fMRI images are used for training and testing the proposed method.

To investigate the relationship between respiratory signal quality and age of the subjects, we divided the respiratory traces into two groups including 306 usable scans (mean age: 15.1 ± 3.5 years) and 2143 not-usable scans (age: 14.5 ± 3.8 years) groups. The difference between the two distributions was statistically significant ($p < 0.05$). In addition to quality of respiratory traces, there is a statistically significant negative correlation between the age of subjects and their head motion ($r = -0.269$, $p < 0.001$). These findings illustrate the challenges of recording respiratory signals during fMRI studies on pediatric populations, where younger participants provide the greatest challenges.

2.3. Data preprocessing

fMRI images were corrected for geometric distortions caused by B0 inhomogeneity using the susceptibility induced distortion correction tool in FSL (TOPUP) (Andersson et al., 2003; Smith et al., 2004); corrected for head motion by registration to the single band reference image using the motion correction tool in FSL (MCFLIRT) (Jenkinson et al., 2002); and registered to the T_1-weighted structural image using Advanced Normalization Tools (ANTs) (Avants et al., 2011). The single band reference image was used as the target image for head motion correction and as the representative fMRI image used to align the fMRI data to the structural data. All the transforms were then concatenated, together with the structural-to-MNI nonlinear warp field registration, and this single resulting warp (per timepoint) was applied to the original timeseries to achieve a single resampling into 2 mm MNI space.

2.4. Proposed method

This study proposes a computational framework for reconstructing full respiratory signal variation waveform from fMRI time series data alone. Due to CNN's unique ability of capturing position and translation invariant patterns, in the proposed method three CNNs are applied in the temporal dimension of the BOLD time series for RV reconstruction. To decrease computational complexity, the average BOLD signal time series from 90 functional regions of interest (ROI) (Shirer et al., 2012) were used as the main inputs to the designed model. These 90 functional ROIs cover most of the cortical and subcortical gray matter. Fig. 2 shows the model input and output. The RV is defined as the standard deviation of the respiratory waveform within a six-second sliding window centered at each time of point (Chang et al., 2009). The output of the model can be defined as any point of the moving window as shown in Fig. 2. In this hybrid method, three approaches are implemented: first point of the window (Method 1 in Fig. 2), middle point of the window (Method 2 in Fig. 2), and end point of the window (Method 3 in Fig. 2). In Method 2, the CNN can use both past and future information, but in Method 1 and 3 it can only use the future and past information, respectively.

Current respiratory pattern impacts the upcoming BOLD signals. To study the influence of breathing pattern on upcoming BOLD signals, the grayscale of BOLD signals is shown by Fig. 3. In this figure, x-axis shows the TRs and y-axis is all voxels in 90 functional ROIs. In this figure, the impact of isolated deep breaths and subject's head motion on BOLD signal is illustrated as an example. In HCP-D dataset, these isolated deep breaths are often followed by a pause in breathing with varying length.

In each of these cases, there are a vertical white band that indicates brief signal increases, and a vertical black band that reflects prominent BOLD signal decreases. This phenomenon is consistent with a reduction in cerebral blood flow following a transitory increase in breathing depth. In addition, the respiratory response function used for modeling a single deep breaths has the same shape (Birn et al., 2008). According to this figure, the variations in breathing pattern induce variation in the BOLD signal and the CNN in Method 1 and Method 2 can use this information to reconstruct the RV timeseries. Current breathing depth and rate are dependent upon previous breaths (Van den Aardweg and Karemaker, 2002). Method 3 can use this information to reconstruct the current RV timepoint.

For Method 2 with windows size 65, as an example, BOLD signals centered at each RV point covering 32 TRs before and after were used as the input. Therefore, each input had a size of [65 × 90], where 65 is the window size and 90 is the number of regions of interest. This broad range for window size (32 TRs before respiration variation point to 32 TRs after respiration) is chosen to encompass both immediate and longer-range respiratory changes. A larger window size provides more information to the model. In the proposed methods, each methods is unable to estimate the RV at the beginning and/or end of the scan, thus a hybrid method to employed to estimate the complete RV: Method 1 estimates the RV at the beginning, Method 2 at the middle, and Method 3 at the end of the scan. Implementing three CNNs with different reference to the BOLD signals enables the RV signal for the whole scan and don't miss the information at the beginning and end of the scan due to effect of sliding window.

No formal rules govern how CNN parameters are set, but the values can be determined by the user based on the problem at hand and their experience. In the five-layer CNN model developed by Salas et al. (2021) the number of filters in different layers are 20, 40, 80, 160, and 320, respectively. In Salas's model, rectified linear unit (ReLU) is used in the hidden layers. This implementation is compared against the proposed model. In the current work, we implement a deep learning CNN with the following architecture: one input layer, six successive one-dimensional (1D) separable convolutional layers, and one output layer. The first input layer represents the input data, which is a tensor with [window size × 90] dimension. The 1D separable convolution layers perform a separate depth-wise convolution on each channel. For the first 1D separable convolution layer in our model, each averaged BOLD signal of a region of interest represents a channel. After applying depth-wise convolution on all channels (90 channels in the first convolution layer), it combines them using pointwise convolution. The number of channels in the first layer can be changed by using other brain atlases. The first convolutional layer has 24 filters of size 3, and the number of filters in the following layers are 32, 64, 128, 160, and 320 respectively with the same size. Activation functions are a critical part of the design of a CNN model. The ReLU is the most common activation function used for hidden layers. A potential disadvantage of ReLU is that it has zero gradient whenever the unit is not active, and may cause units that do not activate initially to never activate as the gradient-based optimization will not adjust their weights. Also, it may slow down the training process due to the constant zero gradients (Gu et al., 2018). To avoid these problems, we used a leaky ReLU that alleviates the 'dead unit' problem with a small slope for negative values. The leaky ReLU compresses the negative part rather than mapping it to constant zero, which allows for a small, non-zero gradient when the unit is not active (Maas et al., 2013). To reduce the size of the input representation by half, a max pooling filter of size 2 is applied to the feature maps after each convolutional layer. Specifically, the pooling reduces the resolution of the extracted features layer-by-layer and simultaneously enhances their robustness. In addition, pooling layers save computation costs and reduce the problem of overfitting. The final layer of the designed CNN model has one neuron with a linear activation function. Fig. 4 shows the architecture of the proposed CNN model.
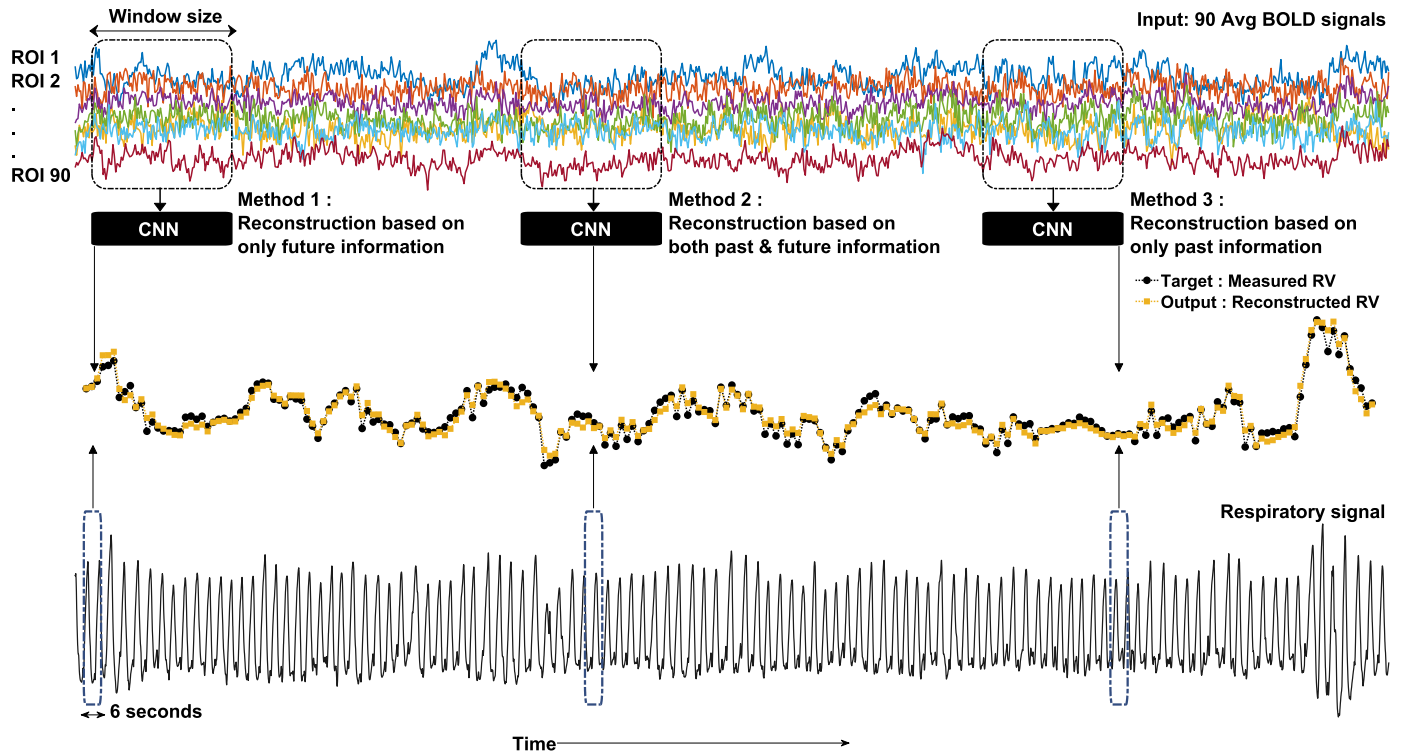
Fig. 2. Inputs and outputs in the proposed method. A respiratory signal is a [152,960 × 1] vector, and RV is a [number of fMRI volumes – 6 second sliding window] vector. Method 1 uses the information in the future BOLD signals, Method 2 takes advantage of both sides, and Method 3 use only the past information. For windows size 65 as an example, Method 1 will estimate $[RV]_{1:32×1}$, Method 2 will estimate $[RV]_{33:446×1}$, and Method 3 will estimate $[RV]_{447:478×1}$. Window size determines the length of the input data and number of averaged BOLD signals in each ROI determines the number of channels.
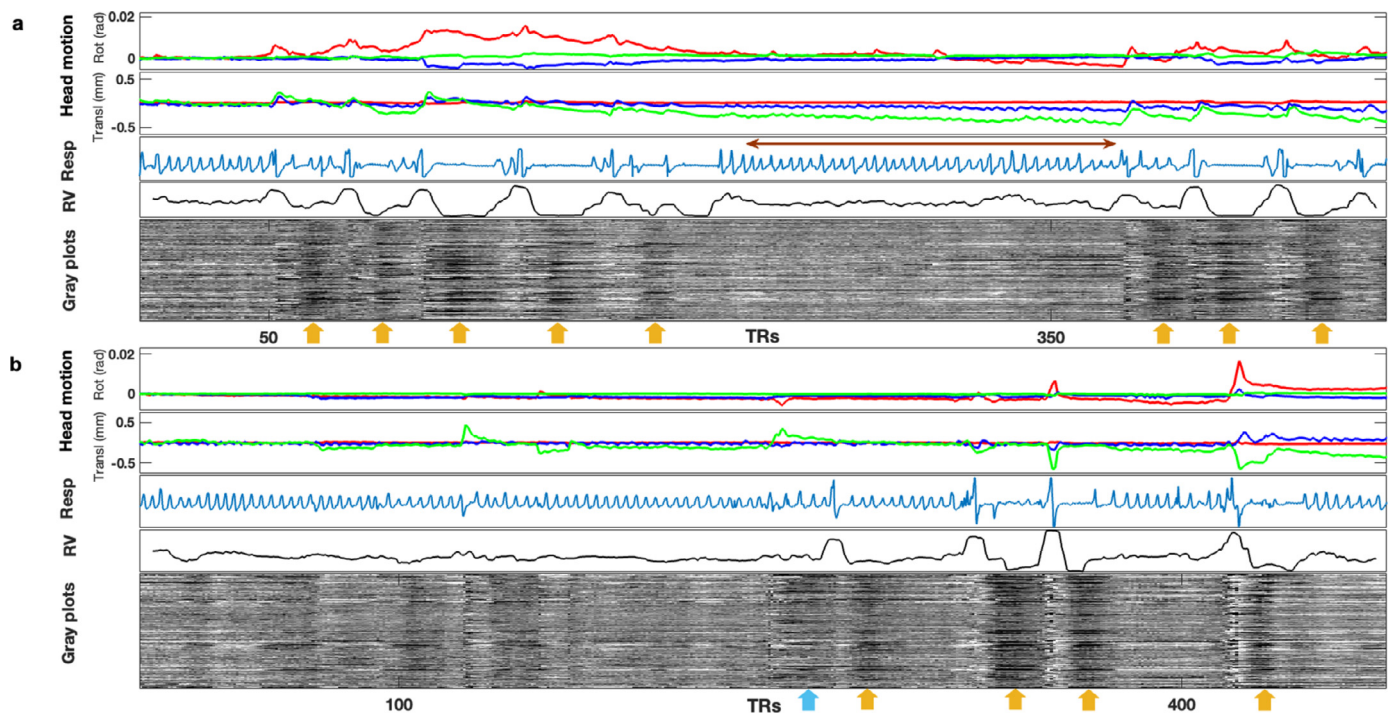
Fig. 3. Illustration of respiratory events impact on BOLD signals. In gray scale heat maps of BOLD signals, white vertical bands show BOLD signal increase and black vertical bands show BOLD signal decrease. Impact of deep breaths on BOLD signal. Vertical black bands after each deep breath reflect the BOLD signal decrease. Prominent respiratory events such as an isolated deep breath impact the RV waveform significantly. In these examples, a bump in RV waveform is created because of each deep breath. On the other hand, RV has a flat shape when subject breathes at almost constant rate and depth, like the time periods shown by yellow double arrows in graph 'a'. For RV reconstruction, CNN can take advantage of substantial fluctuations in BOLD signals.
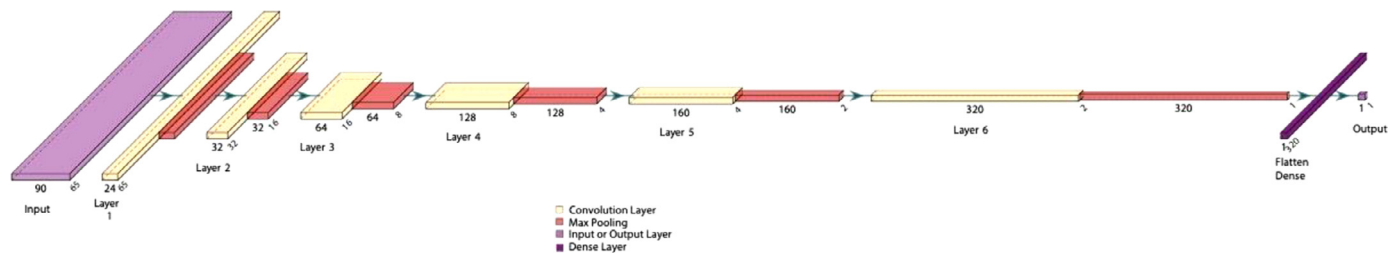
Fig. 4. Architecture of the proposed CNN model. The proposed model consists of six successive 1D separable convolutional layers and one output layer. In the proposed architecture, each convolutional layer is followed by a pooling layer. The first convolutional layer has 24 filters of size 3, and the number of filters in the following layers are 32, 64, 128, 160, and 320 respectively with the same size (tan color). The size of filters in all pooling layers is 2 (red color). This is followed by a flat densely connected layer and then the output layer (dark purple color).

2.5. Evaluation methods

Selection of the proper loss function is critical for training an accurate and robust deep learning model. Certain loss functions will have certain properties and help the model learn in a specific way. Some may put more weight on outliers, others on the majority. The mean absolute error (MAE) is only slightly different in definition from the mean square error (MSE), but provides different results when used in training. Since we are taking the absolute value in the MAE, all the error will be weighted on the same linear scale. Using MAE might result in the model being accurate most of the time but prone to rare, intermittent poor predictions. On the other hand, since the MSE squares the errors, it can help the training algorithm to penalize outlier predictions. As pediatric population are characterized by higher respiration rates and more abnormal breathing patterns, MSE is a more appropriate choice for a loss function. Therefore, we used MSE as the loss function in these experiments.

2.6. Training and evaluation

To train the CNN model and evaluate its performance, respiratory signals in Class 1 and their associated fMRI data are used. Due to limited training data, a ten-fold cross-validation strategy was employed in this work to evaluate the robustness of the proposed model where an independent model is built for each fold there is no model aggregation. Random sampling was performed on the input-output dataset by dividing the data into ten equal portions. For each fold, a model is fit using all samples except the held-out subset (called the first fold). The held-out samples are predicted by this model and used to estimate performance measures. The first subset is returned to the training set and procedure repeats with the second subset held out, and so on. After each run, the trained model is discarded, and a new model trained with the new subsets (Bishop, 2006). At the end, there were 10 independent models. The performance scores from the all 10 runs on held-out samples are then averaged and reported as the "Average ± Standard Error (SE)" (Kuhn and Johnson, 2013). As the held-out or testing subset are kept untouched in each run, there is no chance for data leakage or survivorship during the model training. In each run, the model was trained end-to-end for 300 epochs using a backpropagation algorithm with a batch size of 64. To monitor the training process and prevent the model from overfitting, 20% of the k-fold training data was used to validate the performance at the end of each training epoch. The k value must be chosen carefully, $k = 10$ was chosen as the acceptable trade-off between the higher number of folds and the computational load (Kuhn and Johnson, 2013). To illustrate the performance of the proposed method on unseen data, violin plots are used. In the Results section, each violin plot represents results over the unseen test data in 10-fold cross-validation.

The learning rate (η) of the model was set to 0.001 and was used in conjunction with an Adam optimizer (Kingma and Ba, 2014) to accelerate the learning process. The performance of each fold was evaluated based on MAE, MSE, coefficient of determination of the prediction (R^2),

and dynamic time warping (DTW). The best possible score for R^2 is 1.0. A constant model that always predicts the expected value of y, disregarding the input features, would get an R^2 score of 0.0. In this study, a standard DTW was used with Euclidean distance in the cost matrix is implemented to compare the similarity of reconstructed RV with actual RV. Smaller values of DTW distance indicate a better performance. For interest, the average frequency spectrum of estimated and true RV signals was compared.

The deep network in this work was developed and evaluated in python language using Keras with TensorFlow as the backend. Respiratory signal processing and RV extraction was done using MATLAB version R2021a. Several experiments have been done to evaluate the performance of proposed RV reconstruction method. The computational experiments for this section were done on the Advanced Research Computing (ARC) cluster at the University of Calgary, using two processors with 200 GB RAM to train the network. The code of the proposed method is available on GitHub (https://github.com/jaliladde).

2.7. Ablation studies

Next, ablations studies were performed to determine the effect of the window size and whether the end point or middle point of the window leads to more accurate reconstruction. The window size determines the depth of data that can be changed according to the design priorities such as reconstruction accuracy or maximum length of the reconstructed RV timeseries. In this experiment, the size of moving window changed to values of 9, 17, 33, 65, 129 and 257 to study its impact on CNN's performance. In the CNN models that use smaller window sizes including 33, 17 and 9, the number of pooling layers is decreased to avoid negative dimension size. As an example, the pooling layer after the first convolution layer is removed for a CNN model window size of 33. As changes in breathing pattern have an impact both on past and future BOLD signals, the performance of CNN model for all output definitions, including the moving window's starting point, middle point of moving window and the end point of moving window, is evaluated. Friedman Tests were conducted on different metrics obtained by different settings (for example different window sizes) to examine the effect that parameter had on RV estimation. The hybrid method is then demonstrated, and the method is shown in data without reliably respiration measurements.

3. Results

3.1. Performance of the proposed method

Fig. 5 shows the performance of the proposed method using window size 65 and considering the middle point of the window as the network's output (Method 2) on five samples from the test subset. The trained network can reconstruct the RV timeseries with a high accuracy, especially when there are big changes in RV value. In accordance with Fig. 3, big changes in RV can happen when the subject has a deep breath
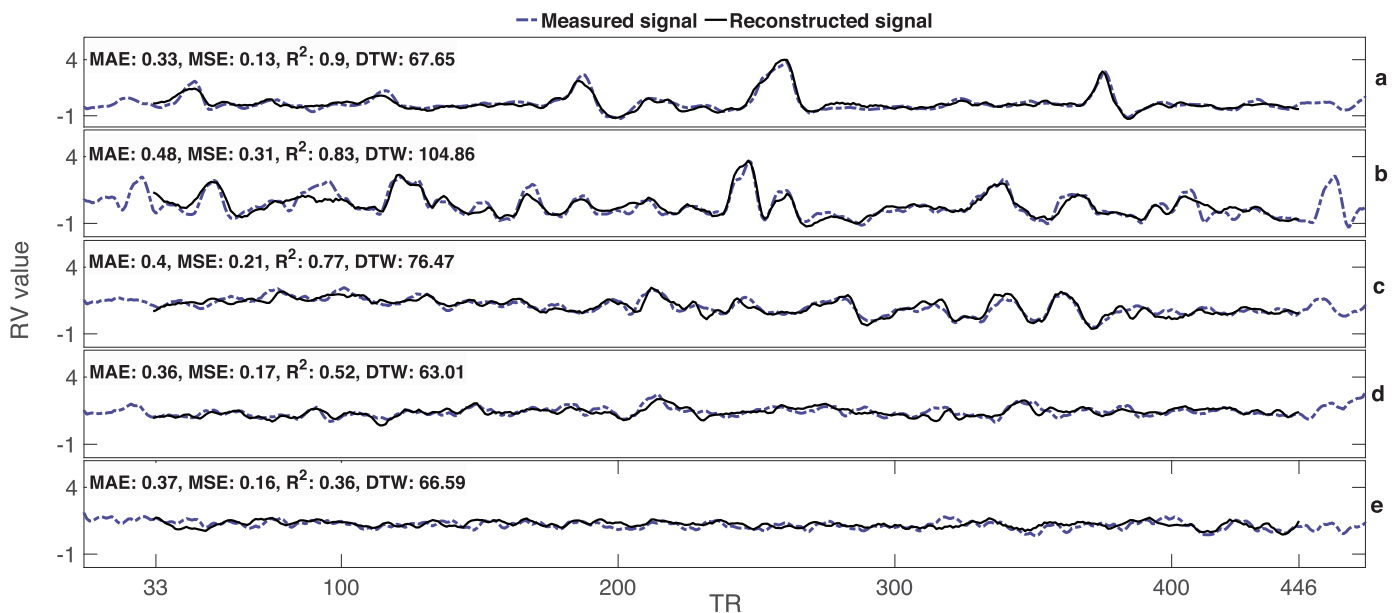
Fig. 5. Performance of the Method 2 with window size 65 on unseen scans. In the HCP-D project, the RV signal length is 478 and using window size 65 in Method 2 leads to loss of 32 points at the beginning and end of the RV timeseries. RV timeseries with higher values and fluctuations are reconstructed with higher correlation performance, and RV timeseries having almost constant values have lower correlation performance but acceptable quality.
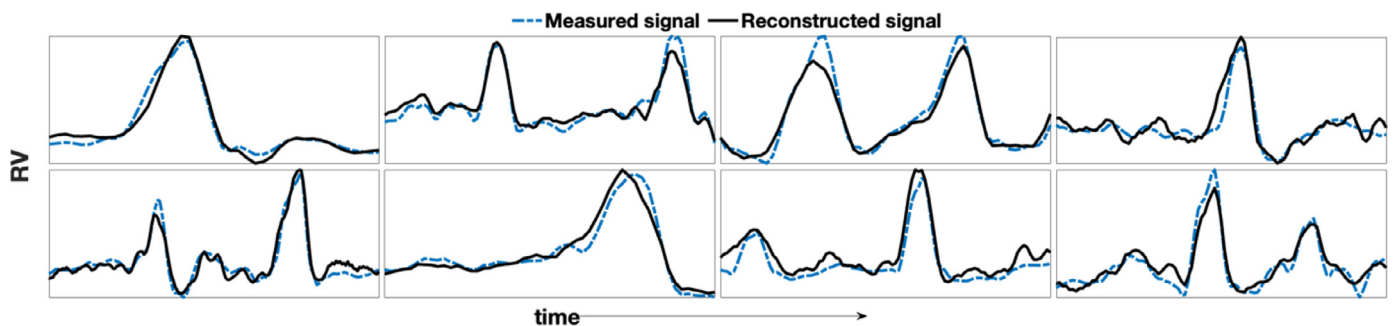

Fig. 6. Performance of the proposed method in reconstruction of RV during deep breaths. There is a high prevalence of deep breaths in fMRI scans. It is essential to capture these events because they cause prolonged and prominent changes in BOLD signals. The proposed method can reconstruct the RV signal during deep breaths with high accuracy.

or changes their breathing pattern. On the other hand, when the subject breaths normally at a constant rate and depth (like the time between two green arrows in Fig. 3a), it is challenging for the CNN to find useful information to reconstruct the RV.

From 352 analyzed respiratory signal in Class 1, 1163 deep breaths (in average, 3.3 deep breath during each fMRI scan) detected. The performance of developed model in reconstruction of RV during deep breaths is evaluated. The developed model can detect the deep breaths from BOLD signal fluctuation in 1131 cases (96.83% precision) and reconstruct the RV with high accuracy. An RV reconstruction of deep breaths is considered a successful case if the real and reconstructed RV had correlation higher than 0.9. Several successful RV reconstructions during deep breaths are shown in Fig. 6. In many subjects there are strong interrelationships between respiration and motion as shown in Fig. 3. But there are some cases that subject has a motion without any respiratory events (shown by a blue arrow in Fig. 3b). In addition, there are some anomalies in the heat maps without obvious causes that could be hardware related (Power, 2017). For CNN training, BOLD signals may be misleading when influenced by head motion or hardware artifacts. There were 37 false positive cases that the CNN detected them as deep breath (a bump in the RV waveform).

As shown by Fig. 5, RV timeseries with higher values and fluctuations are reconstructed with higher accuracy, and RV timeseries hav-
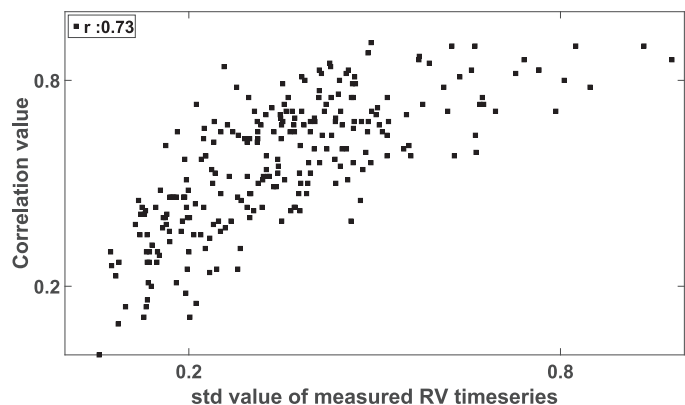

Fig. 7. Relation between RV timeseries changes and CNN's performance. X-axis shows the std value of measured RV timeseries and y-axis shows the correlation value between measured and reconstructed RV. RV timeseries with higher std are reconstructed with higher accuracy.

ing almost constant values are difficult for the CNN to reconstruct. In Fig. 7, the relation between RV timeseries changes in term of standard deviation (std) and the correlation between actual and reconstructed
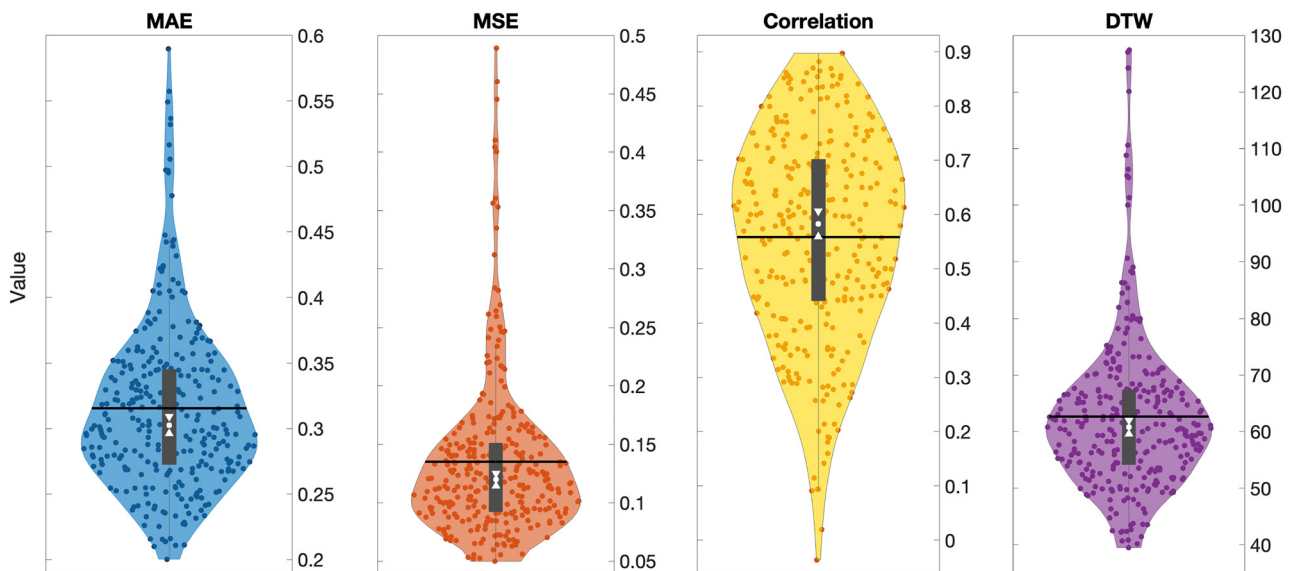
Fig. 8. Performance of the proposed method shown as violin plots of MAE, MSE, correlation, and DTW. Each violin represents results over the unseen test data in 10-fold cross-validation. In these graphs, the white circle and triangle arrows inside the grey box show median and notches, and black horizontal line shows the mean value. Differences among metrics implies that using one loss function for training a machine learning model is not sufficient.

Table 2
Comparing the performance of the proposed method (Method 2) with Salas's model for all scans in the unseen test data in 10-fold cross-validation.

Method	Statistical measure	MAE	MSE	R^2	DTW
Proposed	Mean $\pm$ SE	0.31 $\pm$ 0.0041	0.13 $\pm$ 0.0051	0.558 $\pm$ 0.0098	62.75 $\pm$ 0.91
method	Median $\pm$ SE	0.30 $\pm$ 0.0036	0.12 $\pm$ 0.0029	0.589 $\pm$ 0.012	61.25 $\pm$ 0.61
Salas's	Mean $\pm$ SE	0.34 $\pm$ 0.0064	0.15 $\pm$ 0.0071	0.523 $\pm$ 0.0085	64.37 $\pm$ 1.32
model	Median $\pm$ SE	0.32 $\pm$ 0.0063	0.13 $\pm$ 0.0062	0.532 $\pm$ 0.0135	61.93 $\pm$ 1.08

RV timeseries is shown. It is observed that the CNN can reconstruct the RV timeseries including high variation (higher standard deviation) compared to RV timeseries with low variations.

Fig. 8 shows the performance of the CNN with a window size of 65 and considering the middle point of the window as the CNN's output (Method 2) in terms of MAE, MSE, correlation, and DTW. The performance of the proposed model is compared with the model developed by Salas et al. (2021) in Table 2. More details are available in the supplemental materials. To summarize, the proposed method outperforms the Salas's model with a significant difference between their performance (p-value < 0.01).

Younger individuals breathe at faster rates than adults (20–30 bpm in children between 1 and 5 years-old, 12–20 bpm in children between 6 and 11 years-old versus 10–20 bpm in adults). In Fig. 9, the average frequency spectrum of measured RV and reconstructed RV are shown, which shows similarity of the frequency spectrums. We compared the frequency of all RV timeseries and observed that correlation and DTW lead to higher frequency-based similarity.

3.2. Ablation study

Ablations studies are conducted in many machine learning papers to determine how the model performs as a function of one or more model hyper-parameters. Window size controls the amount of information provided to the model. Fig. 10 shows a sample reconstructed RV timeseries using the proposed method with different window sizes and considering the middle point of the window as the CNN's output (Method 2). The correlation value between the true and reconstructed RV timeseries for window size 9 through 257 are 0.64, 0.66, 0.81, 0.88, 0.93, and 0.96 respectively. Improvement in the correlation value with increas-

ing window size illustrates that changes in breathing pattern have short and long-term effects on BOLD signal, and the CNN can benefit from these long-term effects to improve performance. Larger window sizes by providing more information about baseline breathing rate and depth enable the model to have a better estimate of variation. An important limitation of larger window size is that they result in more timepoints discarded at the beginning and end of the scan due to edge-effects. This factor may limit the minimum duration of scans that can be reasonably reconstructed with the proposed approach, but for longer scans a larger window might be acceptable. For fMRI studies with short scan times, the proposed method using a window size of 17 or 33 will be a reasonable choice with a fair performance.

Fig. 11 shows the performance between actual and reconstructed RV timeseries for the different window sizes. Again, a trend of higher performance is seen with increasing window size. Use of a larger reconstruction time-window could potentially enable a broader range of RV frequencies to be captured by the model, though at the cost of increased complexity. The statistical analyses using Friedman test showed that there is a statistically significant difference between window 17 and 9, 33 and 17, 65 and 33, 129 and 65 (p-value < 0.001), but there wasn't a statistically significant difference between window 257 and 129 for all metrics. Therefore, window size 129 seems reasonable choice for RV reconstruction with highest accuracy.

Fig. 12 shows the impact of output point location, whether we are trying to estimate the point at the beginning, middle or end of the window, on network's performance. The obtained results show that using both side information, before and after the current breath, leads to a better performance (Method 2). There isn't a significant difference between different methods (p-value < 0.001). In general, the performance of all three methods is acceptable.
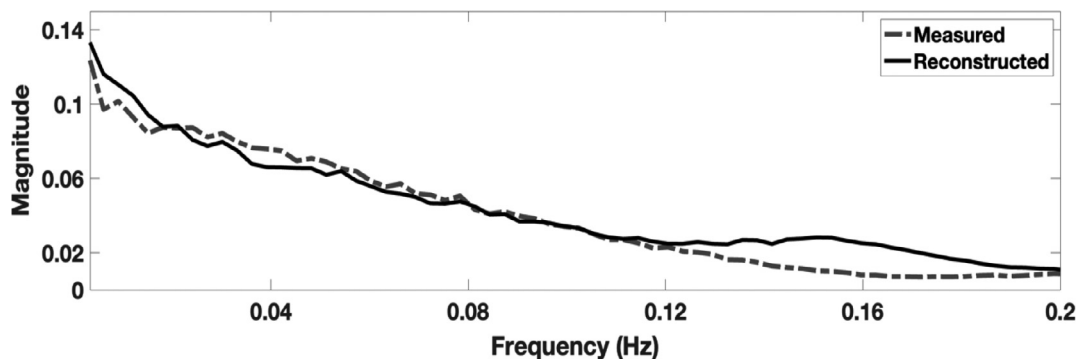
Fig. 9. Average of frequency spectrum of measured RV and reconstructed RV using CNN with window size 65 and considering the middle point as the output (Method 2). As variation in RV timeseries indicate the change in breathing depth and rate, similar frequency spectrum implies the great performance of the proposed method. There is a noticeable mismatch at higher frequencies that shows that the proposed method cannot reconstruct the high-frequency content of RV timeseries. These results indicate the necessity for defining a proper loss function including frequency features.
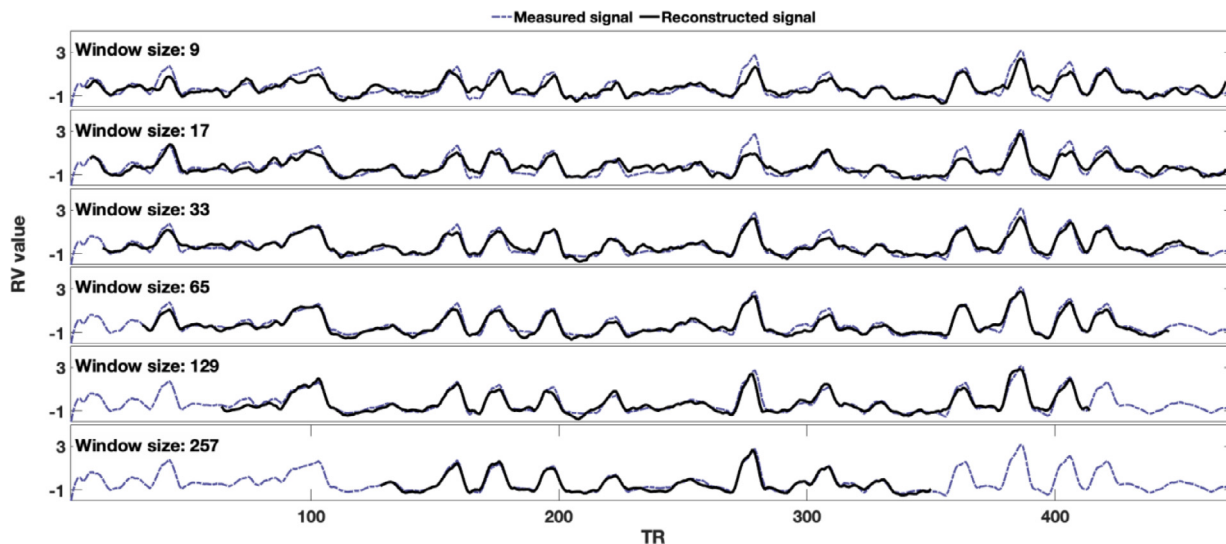
Fig. 10. Impact of window size on the reconstruction accuracy and length of reconstructed signal. By increasing the window size, reconstruction accuracy is improved at the cost of more missing data.

The performance of the hybrid method is shown in Fig. 13. For this experiment a window size of 65 is used, thus 32 timepoints at the beginning is estimated by Method 1, and 32 timepoints at the end are estimated by Method 3. The remaining 414 timepoints at the middle of the signal are estimated by Method 2. Prominent respiratory events such as deep breaths occur at any point of a scan in HCP-D data, but they are more common at the middle and end of the scan. Therefore, the performance of Method 1 in estimation of first 32 timepoints is a little lower compared to its performance on full scan. In addition, there is a small shift in RV value when we switch from Method 1 to Method 2, and form Method 2 to Method 3 as shown by pink arrows in Fig. 13.

An interesting application of the proposed method is reconstruction of the RV time course for scans without respiratory data or scans with poor quality respiratory data. The performance of the developed method is tested on two scans belonging to Class 5 and 6. In this experiment, the full RV signal is reconstructed using Method 1, Method 2, and Method 3. There is a no systematic way to evaluate the accuracy of the reconstructed RV signals in Fig. 14. Therefore, it is compared with the motion and with the BOLD signals in 90 functional ROIs. Considering the vertical bands in heat maps and subject's head motion traces, it can be seen that the proposed method detects motion and signal consistent with deep breathes. For other portions of the timeseries, the BOLD signal and head motions don't indicate any prominent respiratory events (as estimated by the proposed method).

4. Discussion

4.1. Impact on fMRI studies

This work demonstrates the ability to compute the full RV signal from the fMRI data alone in a pediatric population, and we have made notable advances with the approach previously proposed by Salas et al., demonstrated on an adult population (Salas et al., 2021). An interesting potential application of the developed method is enriching a large volume of existing fMRI datasets through retrospective addition of respiratory signal variation information. For example, the Pediatric Imaging, Neurocognition, and Genetics (PING) study (Jernigan et al., 2016) contains fMRI images of 1493 participants (age 3 to 21 years old), where no respiratory signals have been recorded. Cambridge Centre for Ageing and Neuroscience dataset (Shafto et al., 2014; Taylor et al., 2017), IMAGEN study (Mascarell Maričić et al., 2020), and UK Biobank Brain Imaging project (Miller et al., 2016) are some other examples of widely used fMRI datasets where respiratory signals were not recorded. Even in the fMRI studies where respiratory data is collected, a significant proportion of the data is corrupted and not usable. In our analysis, up to 87% of the respiratory signals in the HCP-D project are not usable. Moreover, challenges of respiratory signal measurement can be different in each fMRI site because of different technologies or research assistants. For example, the fMRI data in Adolescent Brain Cognitive Development
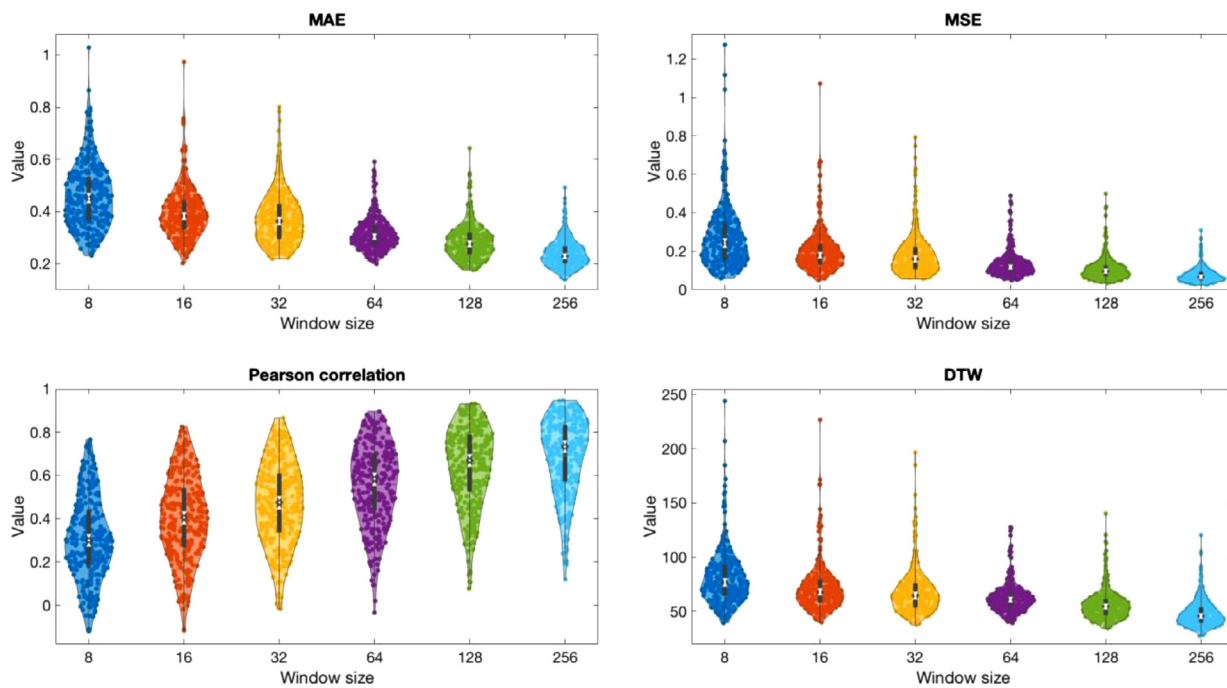
Fig. 11. Impact of window size on reconstruction accuracy in term of MAE, MSE, correlation and DTW values. Comparing the median and mean values shows that using larger windows improve the generalization ability of CNN model.
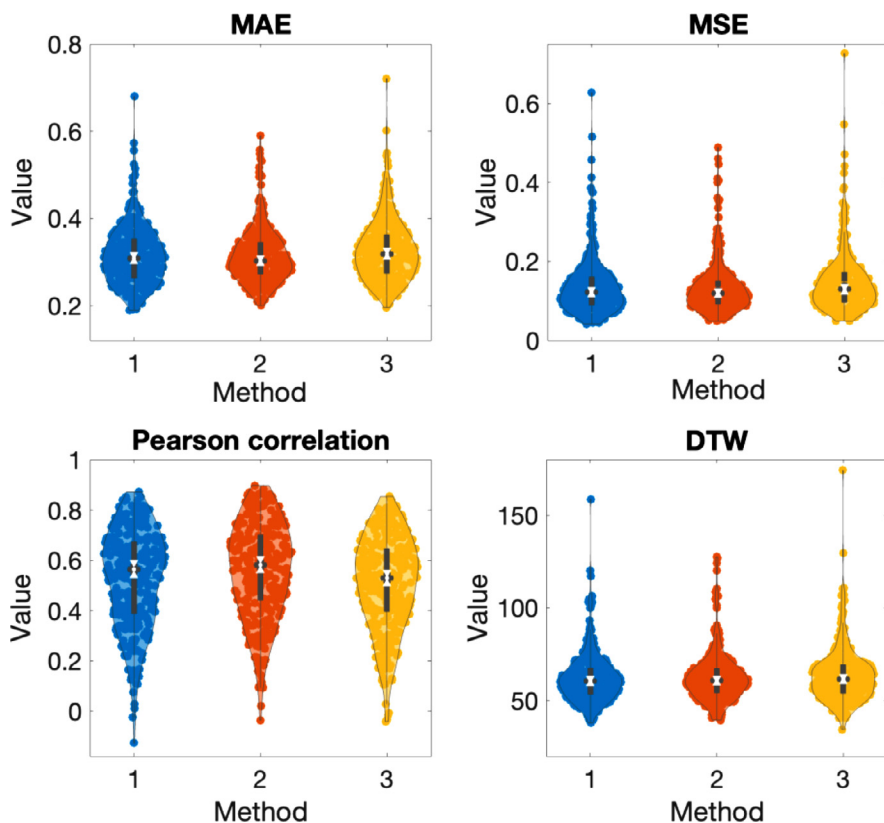

Fig. 12. Impact of output point location, beginning, middle or end point of the moving window, on network's performance with window size 64. The different performance of three methods implies the importance of using input information properly.

study is collected at 21 different sites across the US (Casey et al., 2018) or for the IMAGEN study, across eight different countries in Europe. The method presented here could contribute RV signals to help improve physiological noise compensation.

In addition to removing respiratory effects from fMRI data using the reconstructed RV timeseries by the proposed method, the RV data could provide valuable insight into brain physiology and interpretation of BOLD signals. Higher-order properties of the respiratory signals, such as RV timeseries, might be used to probe autonomic nervous system function and emotional response, which could add depth to the interpretation of fMRI data. Furthermore, the reconstructed RV waveform could provide useful information regarding whether the respiration depth and
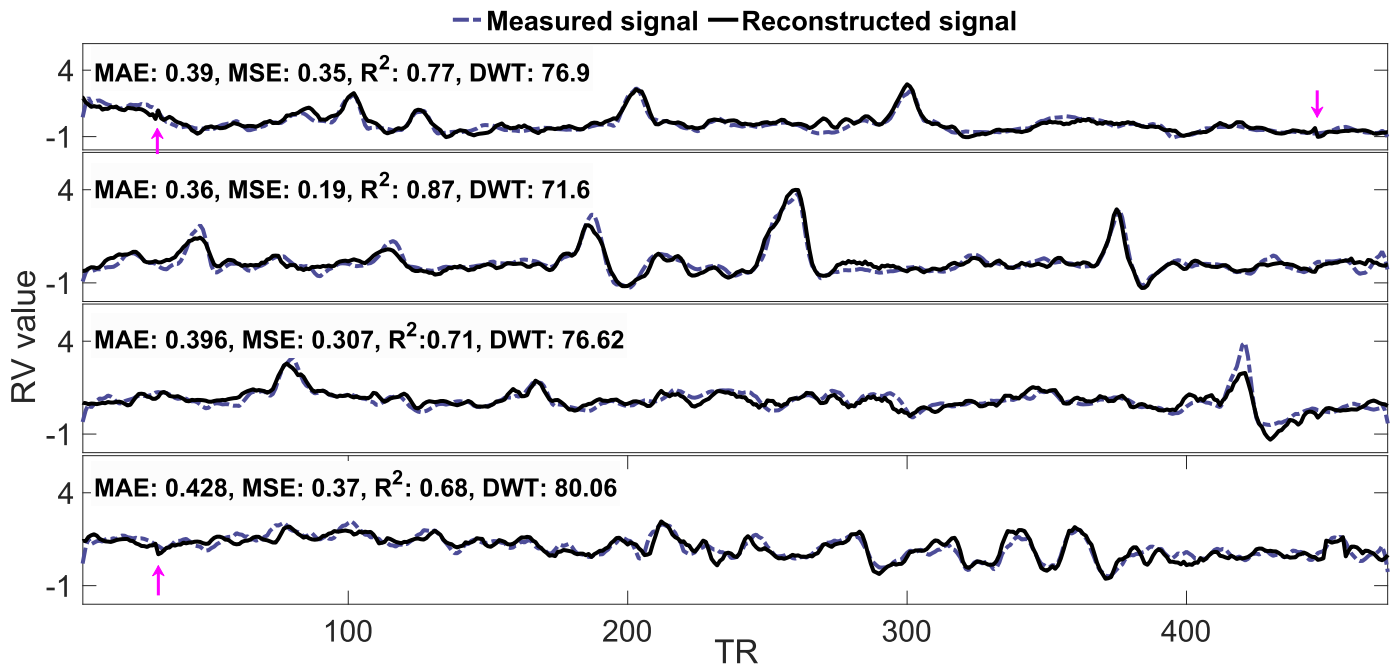
Fig. 13. Performance of the proposed method in reconstruction of whole RV signal. At the switch point between different method, there is a jump in RV value (shown by pink arrows) that indicates the different performance of different methods.
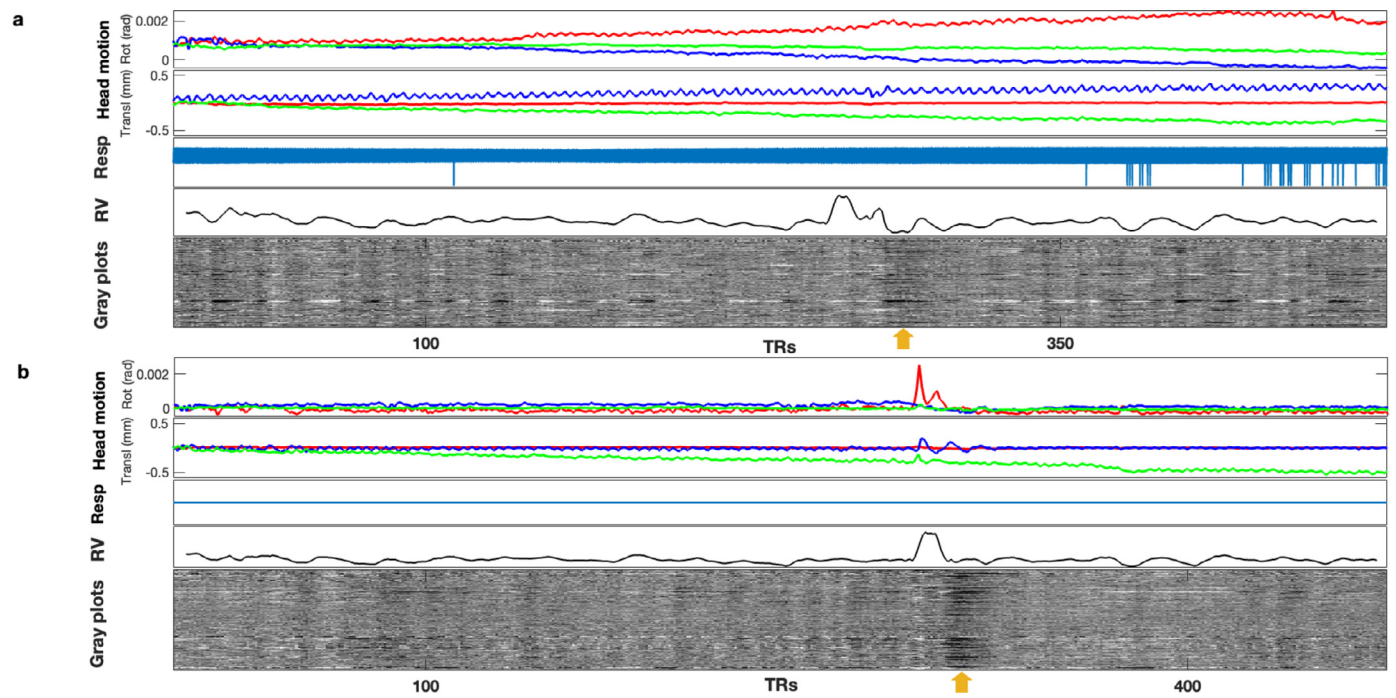

Fig. 14. Performance of the model on subjects with problematic respiratory patterns. A) signal from Class 6. B) signal from Class 5. Vertical black band in the heat maps could be created due to an isolated deep breath. Subjects head motion traces also could be created because of a prominent change in the breathing pattern. According to the information extracted from heat maps and head motion traces, the reconstructed RV appears to be accurate, at least during the deep breath.

rate of a subject differ from scan to scan, which might aid in interpreting fMRI data.

Childhood is a period of rapid brain maturation. Inter-subject consistency has been used to study extrinsic and intrinsic systems in the brain (Hasson et al., 2010; Pajula et al., 2012; Ren et al., 2017). Removing BOLD components induced by changes in breathing rate and depth reduces the inter-subject variability (Birn et al., 2014; Golestani and Chen, 2020). By reducing inter-subject variability via removal of res-

piratory effects, the proposed method could enhance studies of human brain in more realistic and natural settings.

The effect of respiratory timeseries variation is particularly important for resting-state functional connectivity because functional networks are identified on the strength of spontaneous correlations between different regions, and spatially extended artifacts can adversely bias the correlations. It is well-established that the occurrence of neurodegenerative diseases causes changes in the brain resting-state functional con-

nectivity (Izadi-Najafabadi et al., 2022). Even in the early stage of the neurodegenerative disease, it will be difficult for these patients to follow all the guidelines of the imaging process. The proposed method has potential to eliminate the need for additional respiratory timeseries measurement equipment and reduces the complexity. Furthermore, better understanding and quantifying the physiological effects on fMRI studies can accelerate the discovery of connectivity-based biomarkers for diagnosing neurological disorders.

4.2. Comparison with HCP-YA and HCP-A

We evaluated the respiratory signals of 614 subjects in HCP-D (HCD0001305 to HCD2996590), 624 subjects in HCP-YA (115,017 to 350,330), and 550 subjects in HCP-A (HCA6002236 to HCA9007968). The age range for these projects are 5 to 21 years for Developmental, 22 to 35 years for Young Adult (Smith et al., 2013), and 36 to 100 years old for Aging project (Harms et al., 2018). Comparing the scan time of HCP-YA (∼15 min) with HCP-D and HCP-A (∼7 min), more data corruption in HCP-YA project was expected. Using the quality control procedure described in Section 2.2 of this paper, 14.36% of Developmental dataset, 52.1% of Young Adult dataset and 18.9% of Aging dataset were categorized as Class 1 or good quality respiratory signal. Further investigation of the three Class 1 datasets revealed that there is a big difference between the number of removable spikes. The average number of removable spikes in the Developmental, Young Adult and Aging project are 449 ± 229, 51 ± 135, and 423 ± 314 respectively, indicating significantly more data corruption in children and aging population ($p < 0.001$). The amount of respiratory data corruption in these pediatric and aged population datasets reinforces the importance of developing respiratory effect corrections.

4.3. Comparison with other methods

A typical BOLD signal contains respiratory-related components fluctuating around 0.3 Hz (Birn et al., 2006). In comparison with resting state studies where frequencies of interest are typically below 0.1 Hz, this frequency is relatively high. It is likely that respiratory fluctuations will mix with the BOLD fluctuations of interest at the typical TRs used for whole brain imaging (2 s or larger). In this respect, lowpass or notch filters alone cannot remove respiratory-related components without affecting the signal of interest. In HCP-D project where fMRI volumes are acquired with a shorter TR, the respiratory frequency of 1% of subjects is lower than 0.1 Hz which shows high overlap with resting-state BOLD signals. Although, recent finding suggest that BOLD response related to a neural activity can occur at higher frequencies as well (Chen and Glover, 2015; Kalcher, 2014). In addition, filter-based approaches are inadequate for studying non-stationary signals, as the BOLD signal's frequency contents change over time and cannot be defined by a few frequencies. Apart from the main components related to the respiratory cycles, the frequency of changes in respiration rate and depth (∼0.03 Hz) overlaps with the frequencies of fluctuations of true neuronal-related BOLD components (Birn et al., 2008). Since these non-neuronal BOLD components are severely mixed with neuronal-related BOLD components, filter-based methods are not effective for processing them. Considering the high overlap, modeling methods based on semi-quantitative measurements of belt traces such as RV are better approaches to remove respiratory confounds.

Salas *et al.*, proposed a five layers CNN model with ReLU activation function to reconstruct the RV timeseries and tested the model on HCP-YA dataset (Salas et al., 2021). In addition to having higher quality respiratory traces, the HCP-YA provides fMRI images with lower head motion and longer scan time meaning more input data to the CNN. In our model, we increased the number of hidden layers and filters to extract more effective and informative features. Mean $\pm$ SE of correlation values obtained using our models and Salas model were 0.558 ± 0.0098 and 0.523 ± 0.0085, respectively for Method 2. Similar results obtained

for other metrics including MSE, MAE and DWT. During 10 independent runs of the model using ten-fold cross-validation strategy, our model converged to MSE=0.15 for the training dataset after 23 ± 4 epochs, while it was 36 ± 9 epochs for Salas' model. More accurate, faster, and more robust performance of our model show the importance of selection on hyper parameters. Salas et al. (2021) also used a single unit dense network for RV estimation and achieved successful results. We also implemented single unit dense network, but it performed poorly on HCP-D data. It may be because of the pre-processing steps in the Salas implementation, which included temporal smoothing and downsampling prior to training either model, or perhaps the increased motion, faster and more variable respiration in the children. In comparing the proposed CNN with the Salas model, more details are provided in the Supplementary Material.

4.4. Importance of loss function

The loss function plays an important role in the model's performance as well as its architecture, since the loss function determines how well the model fits the training data. Different metrics shows different performances for RV reconstruction (Figs. 8 and 11), and the CNN's performance cannot be accurately represented by using them separately. Let's consider two scenarios.

Scenario 1: An accurate estimation of an RV signal with high amplitude and variations. In this case, the R^2 will have a large value that indicates good performance of CNN. On the other hand, the value of MSE or MAE could be very low (indicating good performance) or high (because the amplitude of the RV signal is high and a small difference between the real and estimated RV can increase their value). Fig. 5b and e is an example of this scenario.

Scenario 2: Estimation with low MSE or MAE. This scenario happens mainly when the RV signal has a low amplitude. In these cases, the value of RV can be high or low, and the low value of MSE or MAE doesn't guarantee a good RV estimation. Fig. 5d and e are example of this situation.

In this study, several performance metrics were used to assess the performance of the CNN. In addition to performance metrics presented, additional performance metrics might also be considered such as the Cosine Similarity (CS). Each metric has subtle differences, for example the R^2 only considers the trend of changes, but CS considers both the trend of changes and distance between two vectors.

To have an accurate comparison between the real and estimated RV and monitor the CNN's performance, a new loss function that uses several metrics could be helpful. One important issue regarding a new hybrid loss function will be the weight of each metric. If we consider a loss function like $\omega_1 MSE + \omega_2 MAE + \omega_3 CS$, the value of weights ($\omega$) should be selected carefully.

4.5. Future work

In this study we used respiratory belt traces measured simultaneously with the fMRI scan. Unfortunately, the redundant information on the abdominal respiratory belt is not backed up by any other source such as a chest belt. It means that it will index both respiratory motions and non-respiratory motions involving abdominal shifts and/or abdominal tightening, but not chest breathing. Therefore several semi-quantitative measurements such as RV, respiratory volume per time (Birn et al., 2006), and envelope of the respiratory trace (Power et al., 2018) have been proposed to index respiratory phenomena from the belt traces. In a study by Golestani and Chen (Golestani and Chen, 2020), effect of end-tidal partial pressure of carbon dioxide ($P_{ET}CO_2$) clamping and retroactive CO_2 correction on functional connectivity and weighted global brain connectivity is compared. Despite expecting similar results from both approaches, they found significant differences between clamped and retroactively corrected connectivity values. In this study we showed that CNN has a great ability to reconstruct the respiratory variation

information from BOLD signals. Therefore, in fMRI studies with large number of participants like UK Biobank, it is reasonable to record the respiratory information of a small portion of the participants using an accurate method such as optical or inductance respiratory plethysmography and design a CNN model to reconstruct the RV signals and remove its effect from the entire cohort.

In addition to other sources of head motion, recent advances in MRI scanning acquisition techniques have revealed that respiration is also an important contributing factor (Fair et al., 2020; Kaplan et al., 2022). While existing motion correction techniques aim to eliminate all motion artifacts, they fail to differentiate respiration-related perturbations that do not disrupt BOLD signals from spontaneous isolated head movements. Head motion due to respiration may lead to an unnecessary reduction in data retention as frame motion estimates selected for elimination may contain a residual respiratory component that needs to be taken into account independently. These artifacts can be removed by applying notch filter, where appropriate age-specific cut-off frequency improved data quality and data retention (Kaplan et al., 2022). In the future work, a CNN model could be trained to determine the main frequency of respiration and design a notch filter customized for each subject.

In this paper, the fMRI data that was used selected as good quality respiratory signals. It is possible that discarded scans may have some specific noise properties that is not present in the scans selected for training. This may result in overtraining to properties of the selected datasets, and potentially create survivorship bias. For example, subject stress may influence the data and we end up discarding scans where participants are stressed. In HCP-D project, subjects were scanned 4 times, two scans per day on two days. Our investigation showed that 35 subjects (5.7% of subjects) have corrupted respiratory signal at day-1, and have clean respiratory signal at day-2, or vice versa. This fact may indicate that the subject's had different emotional state at two different days. Having access to respiratory variation information of these subjects can help the interpretation of the data. Respiratory signals of three subjects with corrupted data at one day and clean data at another day are shown in Supplementary data. In this study we didn't use scans with corrupted respiratory signals, as we needed input-output pairs to train the model. Therefore, our trained model is slightly biased to subjects with clean respiratory behaviour. In the future works, it is worth repeating the fMRI scan until obtaining a clean respiratory signal (for the subject with corrupted respiratory data) and include them in the training/testing dataset to have a model with less bias.

5. Conclusion

Considering the difficulties of respiratory signal recording during fMRI scans in pediatric populations, it is important to understand and develop respiratory effect correction pipelines that address the problems inherent to this population. In this paper we proposed a deep neural network model for reconstruction of full RV timeseries in a pediatric population. The obtained results show that the extraction of respiratory RV timeseries directly from fMRI-BOLD data of pediatric population is possible, especially when the subject has a notable change in their breathing pattern. There exists room to improve the RV timeseries reconstruction accuracy by modifying the architecture of the model, loss function, considering frequency-domain information, subject's head motion traces, and using more accurate respiratory traces as the training data of CNN.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Credit authorship contribution statement

Data availability

Data is available in the public domain through the Human Connectome Project repository. http://www.humanconnectomeproject.org/data/hcp-project/ Training code and model parameters will be made available on https://github.com/jaliladde.

Acknowledgements

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2023.119904.

References

Andersson, J.L., Skare, S., Ashburner, J., 2003. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. Neuroimage 20 (2), 870–888. doi:10.1016/s1053-8119(03)00336-7.

Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C., 2011. A reproducible evaluation of ANTs similarity metric performance in brain image registration. Neuroimage 54 (3), 2033–2044. doi:10.1016/j.neuroimage.2010.09.025.

Bayrak, R.G., Hansen, C.B., Salas, J.A., Ahmed, N., Lyu, I., Huo, Y., Chang, C., 2021. From brain to body: learning low-frequency respiration and cardiac signals from fMRI dynamics. Medical Image Computing and Computer Assisted Intervention – MICCAI, pp. 553–563. doi:10.1007/978-3-030-87234-2_52.

Bayrak, R.G., Salas, J.A., Huo, Y., Chang, C., 2020. A deep pattern recognition approach for inferring respiratory volume fluctuations from fMRI data. Medical Image Computing and Computer Assisted Intervention – MICCAI, pp. 428–436. doi:10.1007/978-3-030-59728-3_42.

Birn, R.M., Cornejo, M.D., Molloy, E.K., Patriat, R., Meier, T.B., Kirk, G.R., Nair, V.A., Meyerand, M.E., Prabhakaran, V., 2014. The influence of physiological noise correction on test-retest reliability of resting-state functional connectivity. Brain Connect. 4 (7), 511–522. doi:10.1089/brain.2014.0284.

Birn, R.M., Diamond, J.B., Smith, M.A., Bandettini, P.A., 2006. Separating respiratory-variation-related fluctuations from neuronal-activity-related fluctuations in fMRI. Neuroimage 31 (4), 1536–1548. doi:10.1016/j.neuroimage.2006.02.048.

Birn, R.M., Smith, M.A., Jones, T.B., Bandettini, P.A., 2008. The respiration response function: the temporal dynamics of fMRI signal fluctuations related to changes in respiration. Neuroimage 40 (2), 644–654. doi:10.1016/j.neuroimage.2007.11.059.

Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer, New York, NY doi:10.1016/j.neuroimage.2007.11.059.

Biswal, B., DeYoe, E.A., Hyde, J.S., 1996. Reduction of physiological fluctuations in fMRI using digital filters. Magn. Reson. Med. 35 (1), 107–113. doi:10.1002/mrm.1910350114.

Casey, B.J., Cannonier, T., Conley, M.I., Cohen, A.O., Barch, D.M., Heitzeg, M.M., Soules, M.E., Teslovich, T., Dellarco, D.V., Garavan, H., Orr, C.A., Wager, T.D., Banich, M.T., Speer, N.K., Sutherland, M.T., Riedel, M.C., Dick, A.S., Bjork, J.M., Thomas, K.M., Dale, A.M., 2018. The adolescent brain cognitive development (ABCD) study: imaging acquisition across 21 sites. Dev. Cogn. Neurosci. 32, 43–54. doi:10.1016/j.dcn.2018.03.001.

Chang, C., Cunningham, J.P., Glover, G.H., 2009. Influence of heart rate on the BOLD signal: the cardiac response function. Neuroimage 44 (3), 857–869. doi:10.1016/j.neuroimage.2008.09.029.

Chang, C., Glover, G.H., 2009. Relationship between respiration, end-tidal CO2, and BOLD signals in resting-state fMRI. Neuroimage 47 (4), 1381–1393. doi:10.1016/j.neuroimage.2009.04.048.

Chen, J.E., Glover, G.H., 2015. BOLD fractional contribution to resting-state functional connectivity above 0.1 Hz. Neuroimage 107, 207–218. doi:10.1016/j.neuroimage.2014.12.012.

Chu, P.P.W., Golestani, A.M., Kwinta, J.B., Khatamian, Y.B., Chen, J.J., 2018. Characterizing the modulation of resting-state fMRI metrics by baseline physiology. Neuroimage 173, 72–87. doi:10.1016/j.neuroimage.2018.02.004.

Chuang, K.H., Chen, J.H., 2001. IMPACT: image-based physiological artifacts estimation and correction technique for functional MRI. Magn. Reson. Med. 46 (2), 344–353. doi:10.1002/mrm.1197.

Fair, D.A., Miranda-Dominguez, O., Snyder, A.Z., Perrone, A., Earl, E.A., Van, A.N., Koller, J.M., Feczko, E., Tisdall, M.D., van der Kouwe, A., Klein, R.L., Mirro, A.E., Hampton, J.M., Adeyemo, B., Laumann, T.O., Gratton, C., Greene, D.J., Schlaggar, B.L., Hagler Jr., D.J., … Dosenbach, N.U.F., 2020. Correction of respiratory artifacts in MRI head motion estimates. Neuroimage 208, 116400. doi:10.1016/j.neuroimage.2019.116400.

Frank, L.R., Buxton, R.B., Wong, E.C., 2001. Estimation of respiration-induced noise fluctuations from undersampled multislice fMRI data. Magn. Reson. Med. 45 (4), 635–644. doi:10.1002/mrm.1086.

Glover, G.H., Lai, S., 1998. Self-navigated spiral fMRI: interleaved versus single-shot. Magn. Reson. Med. 39 (3), 361–368. doi:10.1002/mrm.1910390305.

Glover, G.H., Li, T.Q., Ress, D., 2000. Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. Magn. Reson. Med. 44 (1), 162–167. doi:10.1002/1522-2594(200007)44:1<162::aid-mrm23>3.0.co;2-e.

Golestani, A.M., Chang, C., Kwinta, J.B., Khatamian, Y.B., Jean Chen, J., 2015. Mapping the end-tidal CO2 response function in the resting-state BOLD fMRI signal: spatial specificity, test–retest reliability and effect of fMRI sampling rate. Neuroimage 104, 266–277. doi:10.1016/j.neuroimage.2014.10.031.

Golestani, A.M., Chen, J.J., 2020. Controlling for the effect of arterial-CO2 fluctuations in resting-state fMRI: comparing end-tidal CO2 clamping and retroactive CO2 correction. Neuroimage 216, 116874. doi:10.1016/j.neuroimage.2020.116874.

Golestani, A.M., Kwinta, J.B., Strother, S.C., Khatamian, Y.B., Chen, J.J., 2016. The association between cerebrovascular reactivity and resting-state fMRI functional connectivity in healthy adults: the influence of basal carbon dioxide. Neuroimage 132, 301–313. doi:10.1016/j.neuroimage.2016.02.051.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., Chen, T., 2018. Recent advances in convolutional neural networks. Pattern Recognit. 77, 354–377. doi:10.1016/j.patcog.2017.10.013.

Harms, M.P., Somerville, L.H., Ances, B.M., Andersson, J., Barch, D.M., Bastiani, M., Bookheimer, S.Y., Brown, T.B., Buckner, R.L., Burgess, G.C., Coalson, T.S., Chappell, M.A., Dapretto, M., Douaud, G., Fischl, B., Glasser, M.F., Greve, D.N., Hodge, C., Jamison, K.W., Yacoub, E., 2018. Extending the human connectome project across ages: imaging protocols for the lifespan development and aging projects. Neuroimage 183, 972–984. doi:10.1016/j.neuroimage.2018.09.060.

Hasson, U., Malach, R., Heeger, D., 2010. Reliability of cortical activity during natural stimulation. Trends Cogn. Sci. 14 (1), 40–48. doi:10.1016/j.tics.2009.10.011.

Hu, X., Le, T.H., Parrish, T., Erhard, P., 1995. Retrospective estimation and correction of physiological fluctuation in functional MRI. Magn. Reson. Med. 34 (2), 201–212. doi:10.1002/mrm.1910340211.

Izadi-Najafabadi, S., Rinat, S., Zwicker, J.G., 2022. Brain functional connectivity in children with developmental coordination disorder following rehabilitation intervention. Pediatr. Res. 91 (6), 1459–1468. doi:10.1038/s41390-021-01517-3.

Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17 (2), 825–841. doi:10.1006/nimg.2002.1132.

Jernigan, T.L., Brown, T.T., Hagler, D.J., Akshoomoff, N., Bartsch, H., Newman, E., Thompson, W.K., Bloss, C.S., Murray, S.S., Schork, N., Kennedy, D.N., Kuperman, J.M., McCabe, C., Chung, Y., Libiger, O., Maddox, M., Casey, B.J., Chang, L., Ernst, T.M., Dale, A.M., 2016. The pediatric imaging, neurocognition, and genetics (PING) data repository. Neuroimage 124, 1149–1154. doi:10.1016/j.neuroimage.2015.04.057.

Kalcher, K., Boubela, R.N., Huf, W., Bartova, L., Kronnerwetter, C., et al., 2014. The spectral diversity of resting-state fluctuations in the human brain. PLoS ONE 9 (4). doi:10.1371/journal.pone.0093375, Article 93375.

Kaplan, S., Meyer, D., Miranda-Dominguez, O., Perrone, A., Earl, E., Alexopoulos, D., Barch, D.M., Day, T.K.M., Dust, J., Eggebrecht, A.T., Feczko, E., Kardan, O., Kenley, J.K., Rogers, C.E., Wheelock, M.D., Yacoub, E., Rosenberg, M., Elison, J.T., Fair, D.A., Smyser, C.D., 2022. Filtering respiratory motion artifact from rest-

ing state fMRI data in infant and toddler populations. Neuroimage 247, 118838. doi:10.1016/j.neuroimage.2021.118838.

Kingma, D., & Ba, J. (2014). Adam: a method for stochastic optimization. Proceedings of the 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA.

Kuhn, M., Johnson, K., 2013. Applied Predictive Modeling. Springer, New York, NY doi:10.1007/978-1-4614-6849-3.

Maas, A.L., Hannun, A.Y., & Ng, A.Y. (2013). *Rectifier nonlinearities improve neural network acoustic models* Proceedings of the 30th International Conference on Machine Learning, Atlanta GA USA. https://ai.stanford.edu/~amaas/papers/relu_hybrid_icml2013_final.pdf.

MacDonald, M.E., Berman, A.J.L., Mazerolle, E.L., Williams, R.J., Pike, G.B., 2018. Modeling hyperoxia-induced BOLD signal dynamics to estimate cerebral blood flow, volume and mean transit time. Neuroimage 178, 461–474. doi:10.1016/j.neuroimage.2018.05.066.

MacDonald, M.E., Dolati, P., Mitha, A.P., Wong, J.H., Frayne, R., 2016. Flow and pressure measurements in aneurysms and arteriovenous malformations with phase contrast MR imaging. Magn. Reson. Imaging 34 (9), 1322–1328. doi:10.1016/j.mri.2016.07.007.

MacDonald, M.E., Williams, R.J., Rajashekar, D., Stafford, R.B., Hanganu, A., Sun, H., Berman, A.J.L., McCreary, C.R., Frayne, R., Forkert, N.D., Pike, G.B., 2020. Age-related differences in cerebral blood flow and cortical thickness with an application to age prediction. Neurobiol. Aging 95, 131–142. doi:10.1016/j.neurobiolaging.2020.06.019.

Mascarell Maričić, L., Walter, H., Rosenthal, A., Ripke, S., Quinlan, E.B., Banaschewski, T., Barker, G.J., Bokde, A.L.W., Bromberg, U., Büchel, C., Desrivières, S., Flor, H., Frouin, V., Garavan, H., Itterman, B., Martinot, J.-L., Martinot, M.-L.P., Nees, F., Orfanos, D.P., consortium, I., 2020. The IMAGEN study: a decade of imaging genetics in adolescents. Mol. Psychiatry 25 (11), 2648–2671. doi:10.1038/s41380-020-0822-5.

Miller, K.L., Alfaro-Almagro, F., Bangerter, N.K., Thomas, D.L., Yacoub, E., Xu, J., Bartsch, A.J., Jbabdi, S., Sotiropoulos, S.N., Andersson, J.L.R., Griffanti, L., Douaud, G., Okell, T.W., Weale, P., Dragonu, I., Garratt, S., Hudson, S., Collins, R., Jenkinson, M., Smith, S.M., 2016. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. Nat. Neurosci. 19 (11), 1523–1536. doi:10.1038/nn.4393.

Pajula, J., Kauppi, J.-P., Tohka, J., 2012. Inter-subject correlation in fMRI: method validation against stimulus-model based analysis. PLoS ONE 8 (8), e41196. doi:10.1371/journal.pone.0041196.

Power, J.D., 2017. A simple but useful way to assess fMRI scan qualities. Neuroimage 154, 150–158. doi:10.1016/j.neuroimage.2016.08.009.

Power, J.D., Lynch, C.J., Dubin, M.J., Silver, B.M., Martin, A., Jones, R.M., 2020. Characteristics of respiratory measures in young adults scanned at rest, including systematic changes and "missed" deep breaths. Neuroimage 204, 116234. doi:10.1016/j.neuroimage.2019.116234.

Power, J.D., Plitt, M., Gotts, S.J., Kundu, P., Voon, V., Bandettini, P.A., Martin, A., 2018. Ridding fMRI data of motion-related influences: removal of signals with distinct spatial and physical bases in multiecho data. Proc. Natl. Acad. Sci. 115 (9), E2105–E2114. doi:10.1073/pnas.1720985115.

Rassler, B., Schwerdtfeger, A.R., Schwarz, G., Pfurtscheller, G., 2022. Negative respiratory sinus arrhythmia (nRSA) in the MRI-scanner - a physiological phenomenon observed during elevated anxiety in healthy persons. Physiol. Behav. 245, 113676. doi:10.1016/j.physbeh.2021.113676.

Ren, Y., Nguyen, V.T., Guo, L., Guo, C.C., 2017. Inter-subject functional correlation reveal a hierarchical organization of extrinsic and intrinsic systems in the brain. Sci. Rep. 7 (1), 10876. doi:10.1038/s41598-017-11324-8.

Salas, J.A., Bayrak, R.G., Huo, Y., Chang, C., 2021. Reconstruction of respiratory variation signals from fMRI data. NeuroImage 225, 117459. doi:10.1016/j.neuroimage.2020.117459.

Shafto, M.A., Tyler, L.K., Dixon, M., Taylor, J.R., Rowe, J.B., Cusack, R., Calder, A.J., Marslen-Wilson, W.D., Duncan, J., Dalgleish, T., Henson, R.N., Brayne, C., Matthews, F.E., Cam, C.A.N., 2014. The cambridge centre for ageing and neuroscience (Cam-CAN) study protocol: a cross-sectional, lifespan, multidisciplinary examination of healthy cognitive ageing. BMC Neurol. 14 (1), 204. doi:10.1186/s12883-014-0204-1.

Shechner, T., Wakschlag, N., Britton, J.C., Jarcho, J., Ernst, M., Pine, D.S., 2013. Empirical examination of the potential adverse psychological effects associated with pediatric FMRI scanning. J. Child Adolesc. Psychopharmacol. 23 (5), 357–362. doi:10.1089/cap.2012.0076.

Shirer, W.R., Ryali, S., Rykhlevskaia, E., Menon, V., Greicius, M.D., 2012. Decoding subject-driven cognitive states with whole-brain connectivity patterns. Cereb. Cortex 22 (1), 158–165. doi:10.1093/cercor/bhr099.

Smith, S.M., Beckmann, C.F., Andersson, J., Auerbach, E.J., Bijsterbosch, J., Douaud, G., Duff, E., Feinberg, D.A., Griffanti, L., Harms, M.P., Kelly, M., Laumann, T., Miller, K.L., Moeller, S., Petersen, S., Power, J., Salimi-Khorshidi, G., Snyder, A.Z., Vu, A.T., Glasser, M.F., 2013. Resting-state fMRI in the human connectome project. Neuroimage 80, 144–168. doi:10.1016/j.neuroimage.2013.05.039.

Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M., Matthews, P.M., 2004. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23 (Suppl 1), S208–S219. doi:10.1016/j.neuroimage.2004.07.051.

Somerville, L.H., Bookheimer, S.Y., Buckner, R.L., Burgess, G.C., Curtiss, S.W., Dapretto, M., Elam, J.S., Gaffrey, M.S., Harms, M.P., Hodge, C., Kandala, S., Kastman, E.K., Nichols, T.E., Schlaggar, B.L., Smith, S.M., Thomas, K.M., Yacoub, E., Van

Essen, D.C., Barch, D.M., 2018. The lifespan human connectome project in development: a large-scale study of brain connectivity development in 5–21 year olds. Neuroimage 183, 456–468. doi:10.1016/j.neuroimage.2018.08.050.

Taylor, J.R., Williams, N., Cusack, R., Auer, T., Shafto, M.A., Dixon, M., Tyler, L.K., Cam, C.A.N., Henson, R.N., 2017. The cambridge centre for ageing and neuroscience (Cam-CAN) data repository: structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample. Neuroimage 144, 262–269. doi:10.1016/j.neuroimage.2015.09.018.

Van den Aardweg, J.G., Karemaker, J.M., 2002. Influence of chemoreflexes on respiratory variability in healthy subjects. Am. J. Respir. Crit. Care Med. 165 (8), 1041–1047. doi:10.1164/ajrccm.165.8.2104100.

Williams, R.J., MacDonald, M.E., Mazerolle, E.L., Pike, G.B., 2021. The relationship between cognition and cerebrovascular reactivity: implications for task-based fMRI [Review]. Front Phys 9. doi:10.3389/fphy.2021.645249.

Wise, R.G., Ide, K., Poulin, M.J., Tracey, I., 2004. Resting fluctuations in arterial carbon dioxide induce significant low frequency variations in BOLD signal. Neuroimage 21 (4), 1652–1664. doi:10.1016/j.neuroimage.2003.11.025.

Wowk, B., McIntyre, M.C., Saunders, J.K., 1997. k-Space detection and correction of physiological artifacts in fMRI. Magn. Reson. Med. 38 (6), 1029–1034. doi:10.1002/mrm.1910380625.