Tracing peripheral physiology in low frequency fMRI dynamics

Roza G. Bayrak[1*], Colin B. Hansen[1], Jorge A. Salas[2], Nafis Ahmed[2], Ilwoo Lyu[3], Mara Mather[4,5,6], Yuankai Huo[1,2] and Catie Chang[1,2,7*]

[1]Computer Science, Vanderbilt University, Nashville, 37215, Tennessee, USA.

[2]Electrical and Computer Engineering, Vanderbilt University, Nashville, 37215, Tennessee, USA.

[3]Computer Science and Engineering, Ulsan National Institute of Science and Technology, Ulsan, 44919, Korea.

[4]Gerontology, University of Southern California, Los Angeles, 90089, California, USA.

[5]Psychology, University of Southern California, Los Angeles, 90089, California, USA.

[6]Biomedical Engineering, University of Southern California, Los Angeles, 90089, California, USA.

[7]Biomedical Engineering, Vanderbilt University, Nashville, 37215, Tennessee, USA.

*Corresponding author(s). E-mail(s): roza.g.bayrak@vanderbilt.edu; catie.chang@vanderbilt.edu;

Abstract

Many studies of the human brain using functional magnetic resonance imaging (fMRI) lack physiological measurements, which substantially impacts the interpretation and richness of fMRI studies. Natural fluctuations in autonomic physiology, such as breathing and heart rate, provide windows into critical functions including cognition, emotion, and health, and can heavily influence fMRI signals. Here, we developed *DeepPhysioRecon*, a Long-Short-Term-Memory (LSTM)-based network that decodes continuous variations in respiration amplitude and heart rate directly from whole-brain fMRI dynamics. Through systematic evaluations, we investigate the generalizability of this approach across datasets and experimental conditions. We also demonstrate the importance of including these measures in fMRI analyses. This work highlights the importance of studying brain-body interactions, proposes a tool that may enhance the efficacy of fMRI as a biomarker, and provides widely applicable open-source software.

1 Introduction

The brain and body are closely coupled and continuously influencing each other. Brain-body interactions underpin key functions, including cognition and emotion, as well as the overall health of an organism (Barrett and Simmons, 2015; Shokri-Kojori et al, 2018; Azzalini et al, 2019; Koban et al, 2021).

Functional magnetic resonance imaging (fMRI) is a powerful and widely used technique in human brain research. While fMRI studies do not routinely incorporate physiological signals measured from the body, there is a growing trend toward acquiring continuous physiological measurements (such as heart rate and respiration) during fMRI scans. One motivation stems from considerations about interpretation and reproducibility of fMRI, a pressing issue in the neuroscience field (Botvinik-Nezer et al, 2020). Since fMRI is based on measurements of blood oxygenation, fMRI signals are influenced not only by spatially local changes in neural activity (Ogawa et al, 1992; Kwong et al, 1992; Bandettini et al, 1992), but also by any bodily physiological process that modulate blood oxygenation (Murphy et al, 2013). If not modeled, the presence of these additional effects can complicate the inferences drawn from fMRI as well as drive variability across results, such as in the mapping of large-scale brain networks (Birn et al, 2006; Xifra-Porxas et al, 2021).

Moreover, although physiological effects in fMRI are often regarded merely as confounds, several lines of work indicate that they also provide valuable information. For example, breathing and heart rate exert spatially structured, dynamic influences on cerebral blood oxygenation that closely mirror the spatial structure of core neuronal networks, suggesting a close connection (and potential interactions) between the regulation of blood flow and neuronal responses (Chen et al, 2020; Bright et al, 2020). Indeed, blood-flow responses to heart rate variability have potential to strengthen neural connectivity within networks involved in emotion regulation, highlighting the bi-directional connection between brain and body (Nashiro et al, 2022). Further, removing physiological components of fMRI signals has been shown to reduce test-retest reliability of individual differences in functional connectivity (Dubois and Adolphs, 2016).

Two major physiological drivers of the fMRI signal arise from natural, slowly varying ($<$0.15 Hz) fluctuations in respiration volume (RV) and heart rate (HR). RV and HR are thought to influence fMRI signals through altering blood pressure, autonomic tone, and arterial carbon dioxide (Wise et al, 2004; Birn et al, 2006; Shmueli et al, 2007; Chang et al, 2009; Power et al, 2017; Tong et al, 2019; Picchioni et al, 2022). RV and HR variations are found to account for substantial variance in fMRI signals and to have a larger impact on brain functional connectivity measures compared to physiological effects that are directly synchronized with the breathing and cardiac cycles (Xifra-Porxas et al, 2021). Further, the influence of RV and HR substantially overlaps with neuronally mediated BOLD responses, spanning the same low frequency range (0.01-0.15 Hz) and overlapping with widely distributed functional networks (Birn et al, 2006). Therefore, the ability to precisely identify such physiological components of BOLD is crucial for their use as either ‘noise’ or valuable ‘signal’.

However, it is not always possible to acquire clean external physiological measures during fMRI (Glasser et al, 2018), and many existing datasets lack such measures altogether (ADNI (Jack Jr et al, 2015); UK Biobank (Bycroft et al, 2018); HCP 7T Release (WU-Minn, 2017)). Accordingly, computational approaches have been developed for detecting RV and HR effects in the absence of physiological monitoring. Data-driven methods include global signal regression and ICA, yet currently, these methods cannot unambiguously identify RV and HR components of fMRI signals without the use of recorded physiological waveforms for reference. Recent proof-of-concept studies (Bayrak et al, 2020; Salas et al, 2021) indicated that the RV time course can be directly reconstructed from fMRI signals, but did not examine HR. Since HR has been associated with functional circuits for emotion regulation and correlates with dynamic variation in large-scale functional networks, the ability to reconstruct an HR signal directly from fMRI would contribute an additional rich source of information to fMRI studies. Data-driven techniques have been proposed to infer cardiac phase information directly from fMRI (Beall and Lowe, 2007; Ash et al, 2013; Aslan et al, 2019), from which low-frequency variation in heart rate can be derived. Here we aim to integrate HR estimation in the same framework as RV, with the goals of: (1) improving usability through a single end-to-end model, and (2) to leveraging known covariation between RV and HR to improve HR estimation.

Here, we develop a computational approach for inferring slow changes in respiratory volume and heart rate directly from the fMRI signal. Motivated by the inter-dependence of RV and HR, we propose to jointly learn these signals using a multi-task learning (MTL) architecture, and train models to simultaneously learn RV and HR signals. This work aims to enrich the information content of existing and prospective neuroimaging datasets with missing or corrupted physiological information. An earlier form of this work has been presented (Bayrak et al, 2021).

2 Results

In Section 2.1, we introduce the proposed *DeepPhysioRecon* framework. We then show the agreement between measured and decoded RV and HR signals (Section 2.2), and in Section 2.3, we demonstrate how *DeepPhysioRecon* can enable investigating these low-frequency physiological effects in the absence of measured respiration and cardiac data. In Section 2.4, we assess model generalizability and performance across datasets and experimental conditions. Section 2.5 and 2.6 explore the interpretability of our learning framework. Finally, in Section 2.7 we investigate *DeepPhysioRecon* as a potential tool for denoising RV and HR measures derived from corrupted physiological recordings.

2.1 DeepPhysioRecon Framework

We developed a generalizable deep learning framework to estimate RV and HR from fMRI data. We hypothesized that RV and HR share information that can foster their mutual learning, such that joint learning of RV and HR may enhance model accuracy and generalizability. The proposed network architecture (Figure 1) is composed of a bidirectional Long-Short Term Memory (bi-LSTM) block followed by two linear

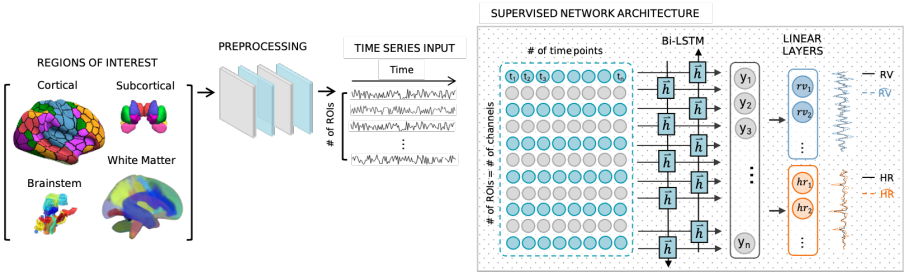


Fig. 1 DeepPhysioRecon Pipeline. The pipeline for estimating respiration volume (RV) and heart rate (HR) signals from fMRI time-series dynamics is shown. Regions of interest are defined using 4 published atlases that had been constructed from different imaging modalities, comprising areas in cerebral cortex, white matter, subcortex, and the ascending arousal network. ROI time-series signals are extracted from the fMRI volumes, detrended, bandpass filtered and downsampled. The preprocessed signals are provided to a candidate network as input channels. A bidirectional LSTM network architecture is adapted for joint estimation. The output of linear layers are RV and HR signals.

layers. An LSTM network allows for learning an explanatory feature space with temporal dependencies, while linear layers hone the focus on each physiological signal separately.

To reduce computational demands and improve signal-to-noise ratio, we employ atlases for dimensionality reduction. As input to the model, fMRI time-series are extracted from 4 different atlases constructed from several imaging modalities, and include cortical, subcortical, white matter, and ascending arousal network regions. The output of the linear layers contains the estimated RV and HR signals, which are of the same length as the input fMRI scans, and sampled at the same rate. The predicted signals are evaluated against the measured RV and HR signals using Pearson correlation.

2.2 Decoding Low-Frequency Peripheral Physiological Signals from Resting State Data

We first demonstrate the applicability of our joint learning approach using the publicly available Human Connectome Project (HCP) resting-state (rs) fMRI dataset. In the HCP rs-fMRI scans, subjects lie quietly with no externally imposed task or stimulation, and each scan lasted for $\sim$ 14.4 min. Cardiac and respiration data are continuously monitored throughout these scans. RV and HR are extracted as the temporal standard deviation of the raw respiration waveform, and the mean beat-to-beat interval, respectively, within a 6-sec sliding window centered on each fMRI time point. Further details of the fMRI and physiological signal preprocessing are described in Section 5.2.

Models were trained using preprocessed HCP rs-fMRI data from 375 subjects scanned 4 times (1500 scans). The resulting models are referred to as *rs-models*. Five-fold cross validation was used. Figure 2a shows the results for one example scan, where decoded signals aligned with the measured (ground truth) signals with a Pearson correlation of r = 0.884 for RV and r = 0.749 for HR. The resulting model accuracy for all folds shown in Figure 2b. The proposed framework reconstructs RV

and HR signals on test cohorts with high agreement against those calculated from the measured physiological data, with a median r $\sim$ 0.689 for RV, and median r $\sim$ 0.627 for HR. To note, the 5 cross-validation folds produced models with highly similar performance on the test partitions (detailed in Section 5.4).

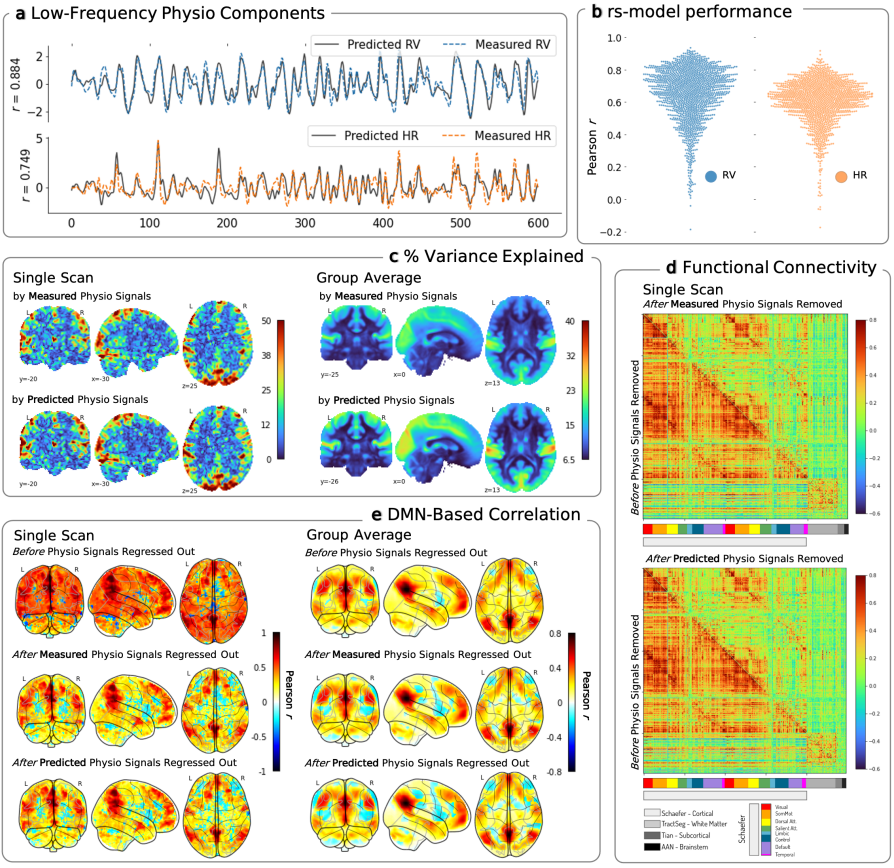


Fig. 2 Resting-state model performance on withheld test data. (a) Low-frequency physiological signals predicted by our model are overlaid on the measured signals for one example scan. Accuracy is measured using Pearson correlation coefficient between measured and predicted signals. (b) Models are trained and evaluated with resting-state data under a 5-fold cross validation paradigm. Each marker represents the Pearson correlation (r) scores between measured and predicted signals for respiration variation (RV – shown in blue) and heart rate (HR – shown in orange), pooled across the withheld resting-state test sets. Single scan and group effects observed on resting-state fMRI data. **Single scan and group effects observed on resting-state fMRI data**. (c) Percent (%) variance explained maps shown for selected slices, indicate the percentage variance explained at each brain voxel by the measured and predicted physiological signals, for one example scan and averaged across the subject group. (d) Functional connectivity matrices highlight the change in ROI-to-ROI correlation after the indicated signals were removed from the ROI time-series data for one example scan. (e) DMN-based correlation maps at the voxel level show the seed-based correlation across the brain regions before and after the measured (as well as predicted) physiological signals are regressed out from the fMRI data, again for one example scan and averaged across the subject group.

2.3 Impact on fMRI signals and network connectivity

We next demonstrate how these decoded physiological signals can improve widely conducted fMRI analyses, and compare the results against analyses conducted using measured ("ground truth") physiological signals. Figure 2 provides examples of the impact of decoding RV and HR at the single-scan and the group levels.

To gauge the influence of RV and HR on fMRI, we first show the proportion of temporal variance explained by a linear combination of RV and HR in the fMRI signal at each voxel, across the whole brain. For this example subject, the variance explained by the predicted RV/HR signals reaches 50% (Figure 2c) – i.e., there are parts of the brain where half of the fMRI signal fluctuation is accounted for by RV/HR. When averaged across the entire set of 1500 scans, the variance explained by RV and HR signals across the brain ranged from 6.5 to 33% (Figure 2c); note that colorbar limits were set to a larger range to maximize image contrast.

We next assess the impact of these decoded physiological signals on functional connectivity analysis. In functional connectivity analysis, correlations between the fMRI time courses of different brain areas are investigated as a proxy for neuronal interactions (Van Den Heuvel and Pol, 2010; Noble et al, 2019; Zhang et al, 2021), but non-neuronal influences (such as RV and HR) can obscure underlying connectivity patterns. When physiological signals are recorded, they are typically removed from the data to control for their influence on functional connectivity. If *DeepPhysioRecon* is successful, regressing out the decoded (versus measured) RV and HR signals should have a similar impact upon functional connectivity.

We therefore assess the functional connectivity between all 497 ROIs drawn from the aforementioned cortical, white matter, subcortical, and ascending arousal network atlases, both before and after projecting out the estimated physiological subspace. Further details of the projection are provided in Section 5.5.1. The Pearson correlation between signals from each pair of ROIs is used to construct a 497 x 497 symmetrical connectivity matrix. As a complementary analysis, we also map the whole-brain, voxel-wise connectivity with respect to a reference ("seed") region in the default mode network (DMN), both before and after removing the estimated physiological subspace. We observe that both the ROI-based (Figure 2d - Single Scan) and seed-based (Figure 2e - Single Scan) correlations are altered with the removal of physiological signals, and that comparable results are obtained using the measured and decoded physiological waveforms, as shown for an example scan. When averaged across all 1500 scans, regressing out the predicted and measured physiological signals had a more mild effect overall but also exerted similar effects on the DMN, enhancing negative correlations while sustaining positive correlations (Figure 2e - Group Average), as expected (Chang and Glover, 2009; Chai et al, 2012). Similar effects of measured and predicted physiological signals on functional connectivity were also obtained at the group level (Supplementary Figure 2).

2.4 Generalizability across tasks and acquisitions

The above results indicate that RV and HR can be decoded from rs-fMRI data. To what degree do these models directly generalize to other fMRI paradigms or acquisition parameters? To test this question, we applied a randomly selected model from the 5-fold CV on resting-state data, without any additional training or fine-tuning, to data acquired during seven different task paradigms from the HCP dataset. Further, as an even stronger test of model generalizability, we also evaluated the performance of the HCP resting-state model, again without any additional training, on an external dataset that was acquired on a different scanner and with different acquisition parameters, including temporal resolution (results from models built from each of the 5 folds are reported in Supplementary Figures 3 and 4). Further details about the datasets and preprocessing are described in Section 5.1, and Section 5.2.

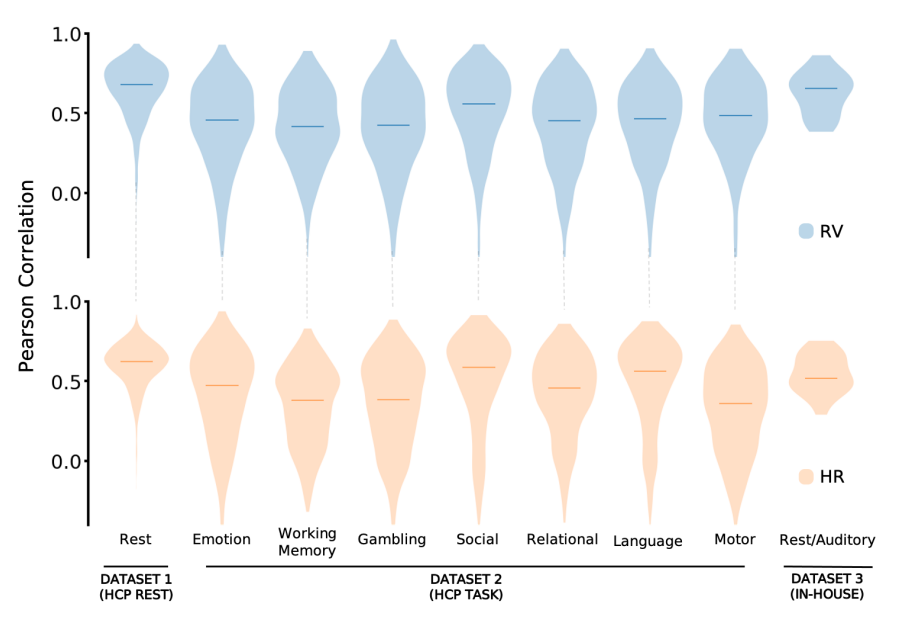


Fig. 3 Generalizability of rs-model to scans acquired under different experimental conditions and acquisition parameters. From the five models trained on resting-state data under the 5-fold cross validation paradigm, a model is randomly selected for this assessment. Seven tasks from Dataset 2 (HCP: emotion, working memory, gambling, social, relational, language and motor), and Dataset 3 (in-house: rest/auditory) were used to assess generalization of the model. Each plot represents the aggregated Pearson correlation scores between measured and predicted signals for respiration variation (RV – shown in blue) and heart rate (HR – shown in orange) in the withheld test set. Median r is indicated by the horizontal line.

The results (see Figure 3) suggest that for the held-out datasets, the model trained on the HCP resting-state data captures a transferable feature space. The agreement between measured and predicted signals for RV was high in the in-house dataset (median $r > 0.6$), and moderate for all HCP tasks. For HR, the performance was also high for the in-house dataset. Moreover, HR had notably high performance on the

HCP social and language tasks (median r $\sim$ 0.6), and a moderate agreement for the other HCP tasks.

Finally, we asked whether training models on task data could improve the performance on datasets involving tasks that are different from those on which the model is trained. This question also probes whether the relationship between fMRI and physiology is largely independent of task condition. To investigate this question, we preserve the same framework used for the rs-model (including hyperparameter values and model architecture) and train a new model, which we will refer to as the **task-model**, using only four of the seven tasks (working memory, social cognition, emotion processing, and gambling tasks). The task-model is then deployed to test generalizability on the three remaining tasks (relational processing, language, motor tasks), as well as on the HCP resting-state scans and the in-house dataset. The results show (see Figure 4) that overall, the task-model is able to learn both RV and HR with moderate agreement against ground truth (median r $\sim$ 0.5). When tested on the HCP tasks, the task-model showed an improvement in performance compared to the rs-model (Supplementary Table 5).

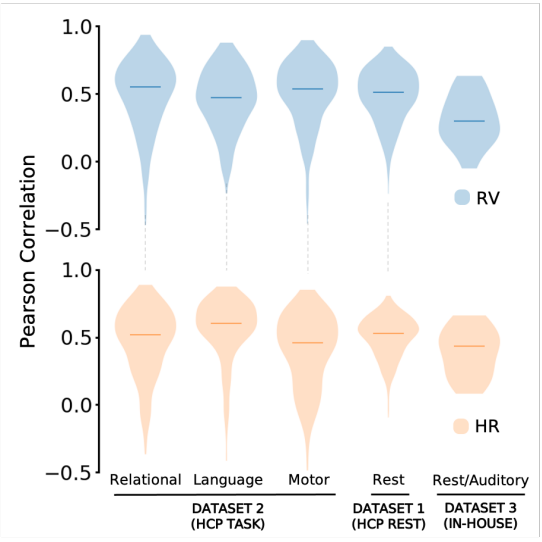


Fig. 4 Generalizability of task-model on test cohorts acquired with different task conditions and acquisition parameters. Three tasks from Dataset 2 (HCP: relational, language and motor), Dataset 1 (HCP: resting-state) and Dataset 3 (in-house: rest/auditory) were used to assess generalization of a model trained on data consisting of 4 other task conditions. Each plot represents the aggregated Pearson correlation scores between measured and predicted signals for all scans in a set. Median r is indicated by the horizontal line.

2.5 Spatially constraining what the network sees

Our learning framework can flexibly accommodate any set of atlas regions for dimensionality reduction. Although we chose a large number of brain regions when fitting

the original model in order to maximize brain coverage, we hypothesized that certain brain regions convey more physiological information than others. For example, regions near major blood vessels may be more strongly influenced by physiology, and thus provide greater predictive value. To investigate this hypothesis, three linked experiments are designed. In the first, we use only one input ROI times-series signal at a time, generating what we will refer to as **individual ROI models**. These models are trained in a manner identical to that of the original (multi-ROI) model. The model performance for each scan is assessed using Pearson correlation between the predicted and measured physiological signals, and an average score for the entire test cohort is noted (see Section 5.4). Figure 5a shows these average scores mapped onto their corresponding regions for all 497 regions, representing the predictive ability of each individual ROI. The best performance of individual ROI models reaches $r \sim 0.5$ (note that colorbar limits were set to a smaller range to maximize image contrast).

Next, we assess whether the number of regions included in training affected the model performance. Here, models are trained using successively increasing numbers of ROIs (from 1% to 100%), rank-ordered by the results of the individual ROI analysis. These models will be referred to as **percent ROI models**. While the improvement in model performance continuously increased with the number of included ROIs (Figure 5b, i.e. for RV median Pearson $r \sim 0.45$ and $r \sim 0.65$ respectively with 1% and 25% of ROIs), the rate of increase slows after the model includes approximately 25% of the top ROIs, and reaches median Pearson $r \sim 0.7$ with 100% of the regions.

Given that brainstem and cerebellar regions are often absent from fMRI acquisitions, and yet are implicated in cardiovascular regulation, the last experiment assesses the performance when using all regions except those that spatially overlap with brainstem and cerebellum. Excluding brainstem and cerebellar regions resulted in comparable performance (Figure 5c).

2.6 Factors driving model performance

We further investigated the various factors that may drive or affect model performance. For this, we first examine whether the results are driven by BOLD signal in regions near blood vessels. To investigate this relationship between the "vesselness" of a region and its relative predictive ability, we carried out linear regression analysis between Time of Flight (ToF) and Susceptibility Weighting Imaging (SWI) measurements of vessel density (Bernier et al, 2018) in an ROI and the corresponding individual ROI model prediction accuracy. The results are shown in (Figure 6a). Interestingly, the vessel density metric is negatively associated with predictive value, showing that the higher the vessel density, the less accurate the model predictions from the fMRI signal within that region.

Another candidate factor could relate to the magnitude of RV and HR fluctuations exhibited by the subject in a given scan, as larger physiological variations would modulate brain hemodynamics more strongly. To investigate the relationship between the magnitude of fMRI physiological responses and model accuracy, we carried out a linear regression analysis between the amount of fMRI temporal variance explained in a scan (averaged across all voxels in the brain) and the model prediction accuracy.

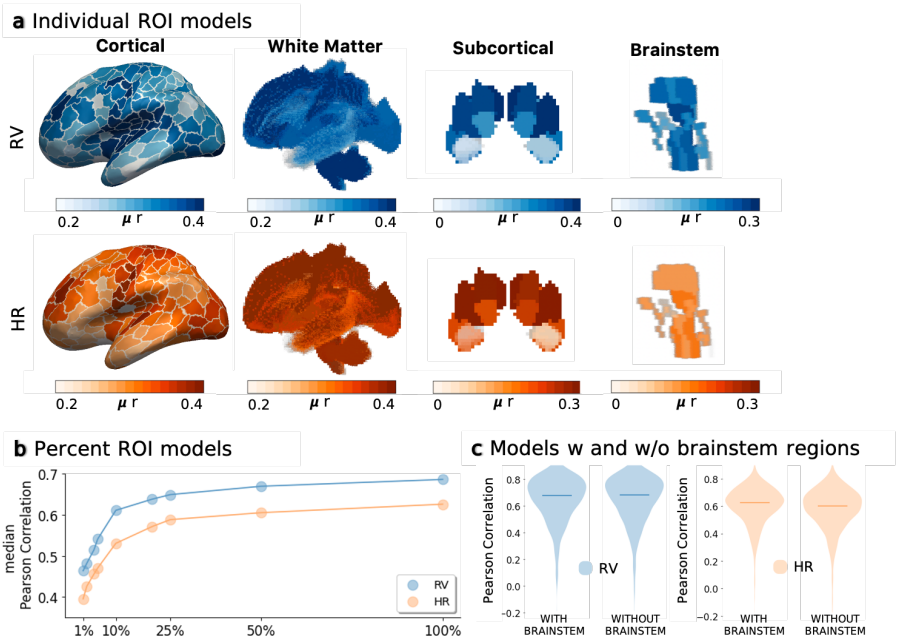


Fig. 5 (a) **Individual ROI models**. Using resting-state data, models were trained separately for each individual ROI. Predictiveness of a given ROI was assessed using Pearson correlation between measured and predicted signals. Results were visualized by projecting the mean r onto (left to right) cerebral cortex (400 ROIs), white matter bundle regions (72 ROIs), subcortical regions (16 ROIs) and ascending arousal network (9 ROIs). (b) **Percent ROI models**. A set of models are trained with an increasing number of ROIs. Starting from the top 1% of the total ROIs, rank-ordered by results of the individual ROI analysis, models are trained using successively increasing numbers of ROIs, up to 100% of ROIs. (c) **Model comparison without brainstem and cerebellar regions**. A model is trained using 477 ROIs, excluding regions that spatially overlap with brainstem and cerebellum. Each plot represents the Pearson correlation scores between measured and predicted signals for all scans in the unseen test data.

The results (Figure 6b) indeed indicate a positive correlation between the two, suggesting that the amount of physiologically induced BOLD variation during a scan impacts the ability to predict RV/HR from an fMRI scan.

2.7 Potential for denoising physiological waveforms

Low correlation scores can emerge when there are artifacts in the measured physiological recordings, coloring the performance during testing and therefore necessitating close inspection. We speculate that the signals predicted from *DeepPhysioRecon* may have the potential to 'fix' some of the artifacts in the measured physiological data. In Figure 7, we provide one example in which possible motion-induced artifacts can be observed in the raw cardiac signal, whose effects propagate to the "ground truth" low-frequency HR signal. We observe that the predicted HR smooths over these artifacts, suggesting that the model may be able to clean up RV/HR measures through this data-driven learning process. Another observation is that in some cases

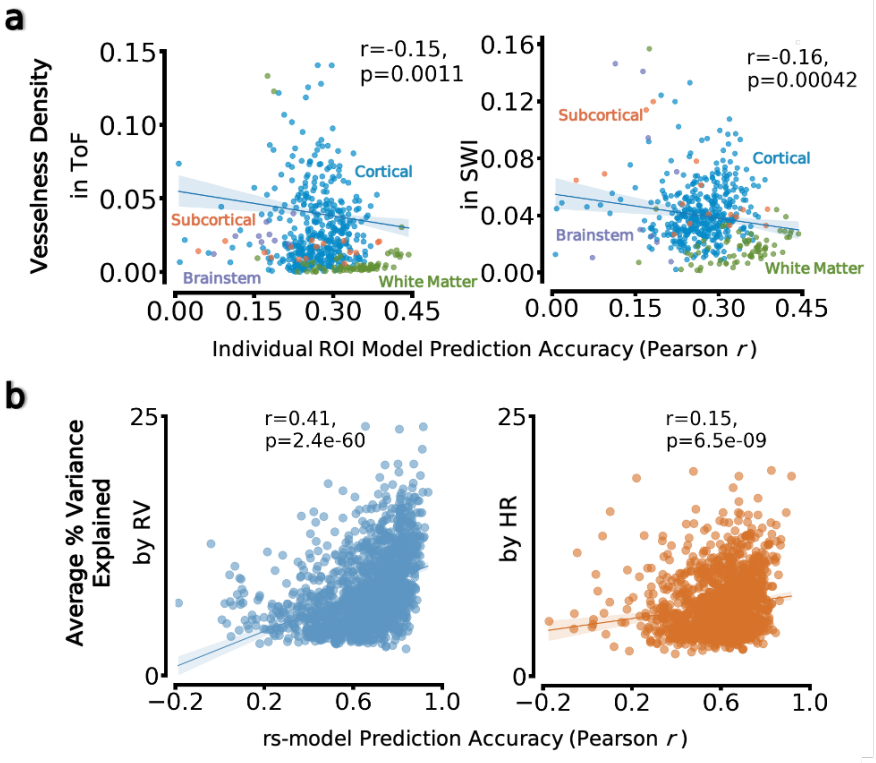


Fig. 6 (a) **Vessel density as a factor that may underlie regional prediction accuracy.** In each plot, the x-axis represents model accuracy for each individual-ROI model. The y-axis represents the average vessel density for Time-of-Flight (ToF - shown in green) and Susceptibility Weighting Imaging (SWI - shown in purple) measures, averaged across all voxels within each ROI. (b) **Percent variance accounted for by physiology in fMRI data, as a factor that may underlie model prediction accuracy.** In each plot, the x-axis represents cross-validated model accuracy calculated using Pearson correlation scores between measured and predicted signals for respiration variation (RV – shown in blue) and heart rate (HR – shown in orange) in the withheld resting-state test set. The y-axis represents the percentage of temporal variance explained in the fMRI data, averaged across all brain regions, by the respective (measured) physiological signal.

wherein the predicted RV/HR signals exhibited low correlation with the ground-truth RV/HR signals, the predicted signals explained a larger variance in the fMRI data compared to the measured physiological signals. Two such examples are shown in Figure 8 (Subject A, Subject B). Notably, although having stronger magnitudes, the spatial maps associated with the predicted physiological signals retain subject-specific patterns that are also present in the maps associated with the measured physiological signals. These examples may also indicate that the model can generate physiological estimates that are cleaner than the recordings. In cases where the measured physiological signals appear to be clean and have high correlation with the

predicted waveforms (Subject C), close correspondence is also seen in their effects on fMRI data.

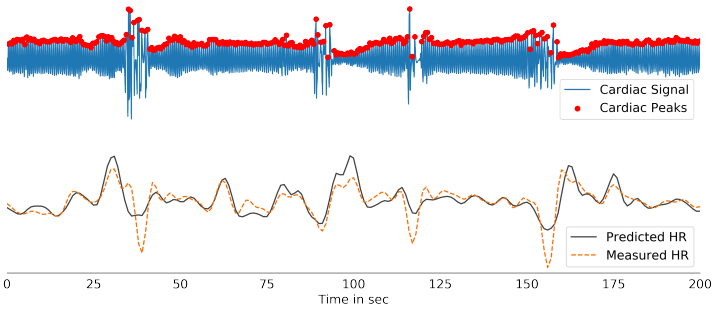
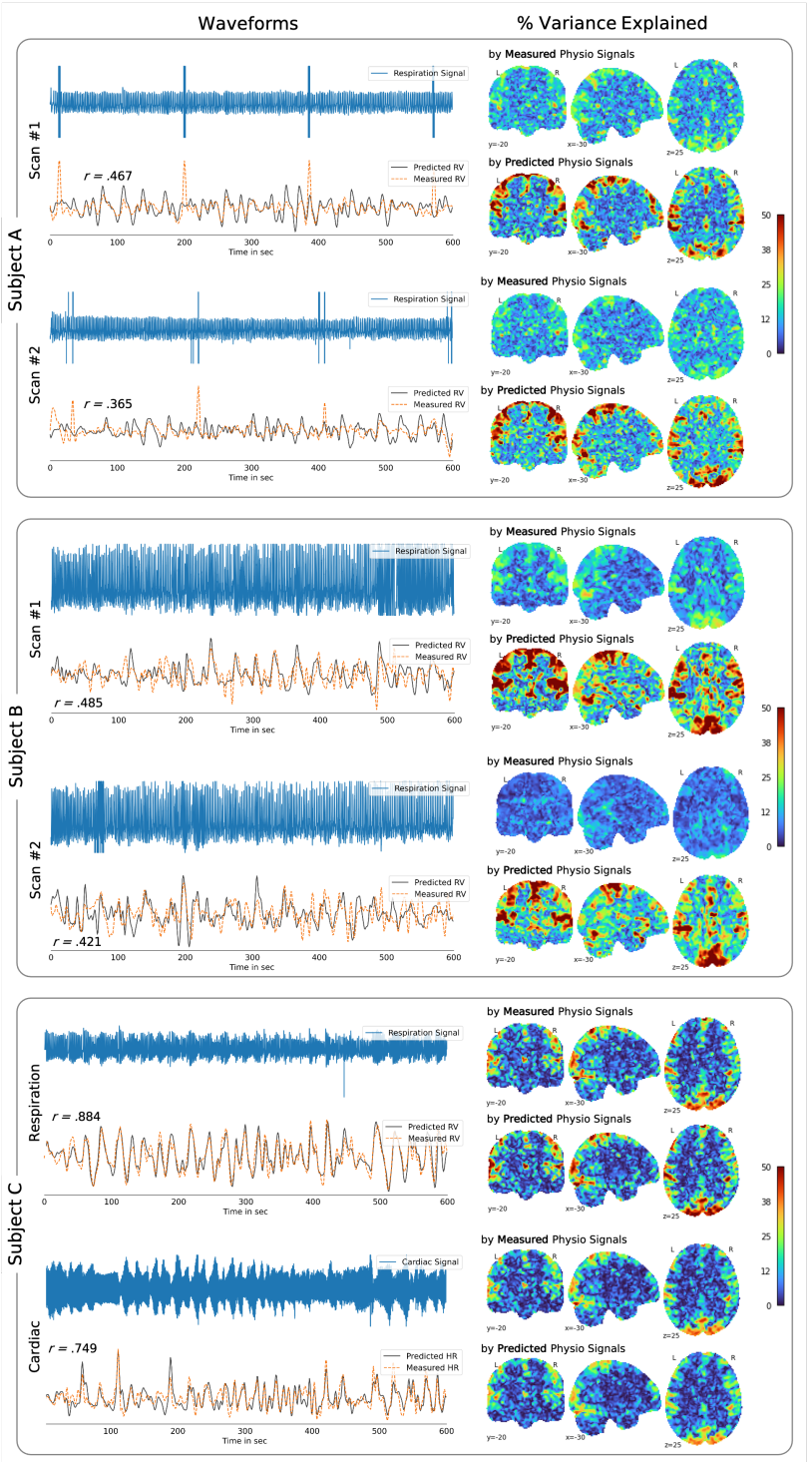


Fig. 7 An example recording is shown. The transient artifacts in the HR waveform (bottom; orange) are likely due to motion in the raw PPG signal (top). DeepPhysioRecon may help to 'fix' noisy physiological recordings, as indicated by the predicted HR signal (bottom; black).

3 Discussion

This work presents a framework for jointly inferring respiration and heart rate fluctuations directly from fMRI dynamics. The *DeepPhysioRecon* framework provides these two key low-frequency physiological signals (RV, HR) to datasets that fully or partially lack external physiological measures, or which lack measurements of sufficient quality. We demonstrate that the proposed models of RV and HR, trained on resting-state fMRI data, generalize across datasets with varying experimental conditions and significant acquisition differences. Consistent with the literature (Birn et al, 2006; Chang et al, 2009; Power et al, 2017; Kassinopoulos and Mitsis, 2019), we also find that RV and HR can, in certain scans, account for large amounts of temporal variance in fMRI signals. By including the predicted RV and HR signals in functional connectivity analyses, we also show that large-scale network maps are altered in agreement with the changes introduced by incorporating the measured RV and HR signals. The proposed models are tested not only in independent participants within each dataset, but also across datasets acquired with different experimental conditions, MR scanners, and protocols.

Models trained and tested on resting-state fMRI data succeeded in decoding RV and HR signals with high accuracy. These models were also found to generalize well between experimental conditions, different subjects, and even across significantly different acquisition parameters, including different temporal resolutions (TR=0.72s to 2.1s). Figure 3 indicates that models that were directly applied to different fMRI conditions (i.e. resting-state models applied to task, and task models applied to resting-state or unseen task conditions), could infer information in a new set of experimental conditions. This ability to transfer models between fMRI conditions suggests that the relationship between fMRI and physiology could be largely brain state-invariant. However, as expected, models trained using task fMRI data exhibited



(Caption on next page.)

Fig. 8 Low correlation between measured ('ground truth') and predicted RV signals may not accurately reflect the quality of reconstruction. In the selected examples, artifacts are observed in the raw time-series signals. For example, data from Subject A is shown to have periodic artifacts likely related to the equipment, and data from Subject B exhibits clipping artifacts. These are then carried over to the derived low-frequency RV and HR waveforms. We observe that the predicted waveforms appear to mitigate some of these artifacts, suggesting that the model may be able to clean up physiological signal measures through this data-driven learning process. The percent variance explained maps indicate that the predicted RV signals accounted for a much larger proportion of the fMRI signal variations compared to the measured RV signals. On the contrary, in the example of Subject C, the respiration and cardiac recordings from the same scan (without any major artifacts) are shown. The variance explained maps indicate that measured and predicted RV signals accounted not only for a similar proportion of the fMRI signal variations but spatial distribution of these maps closely aligned.

improvements on decoding task data in comparison to those trained on resting-state data (Figure 4; Supplementary Table 5), though the degree of performance improvement varied across tasks. Likewise, models trained on resting-state data performed better on held-out resting-state data than they did on task data.

One possible explanation for this increase in accuracy is that within a particular condition (task vs. rest), it is expected that the distribution (signal and/or noise) matches more closely. For the case of task, there may be task-related brain activity that shows up in more systematic ways both in the fMRI data and physiological recordings. Therefore, we can speculate that the model learns to isolate these patterns when trained with task data. Further, the success of rs-models to more accurately decode an external dataset with different acquisition parameters, compared to the t-model, may also relate to the conditions used in the external (in-house) dataset. The in-house dataset tasks involved only simple sensory stimulation (auditory tones) that were presented at long-inter-stimulus intervals (ISI) for some scans, or at very short ISI but continuously throughout the experiment for others. Both of these designs will have more contribution from spontaneous fluctuations, resembling the resting-state conditions.

Nonetheless, we found that even when training and testing on the respective conditions, model performance was higher on the rs-fMRI data than for task fMRI data. One explanation may be that the resting-state scans were longer than the task scans. It has been shown that increasing the scan duration may increase the reliability of fMRI connectivity estimates (Birn et al, 2014), so perhaps training models with long duration ($\sim$ 15 min) scans (as in the HCP resting-state data) may offer advantages over training with the shorter-duration (2-5 min) scans.

We suggest that depending on the goal of the study, different training datasets should be considered. If the goal is to create the data in its absence (i.e., to reconstruct physiological signals in a dataset that does not include any physiological measurements), a large cohort including variety of fMRI conditions and acquisition parameters may give a more robust model. Training models using the same task as the one for which prediction is needed can further increase accuracy, as the models may be trained using the same dataset and same condition. While this will reduce the generalizability of a model across conditions, it may allow for better condition-specific reconstruction as well as for 'fixing' corrupted data within a dataset (i.e., cases in which the dataset has physiological measurements but contains corrupted samples). However in the latter case, we would extra caution against overfitting to the

data, which can be achieved in various ways. For instance, initializing the weights by a model pretrained with large-scale fMRI data across a broad set of conditions can provide a good starting point.

Our neural network architecture was designed to handle many unique challenges related to fMRI. First, bi-directional LSTM networks support varying-length input and multivariate output. This allows for learning on varying-length input signals and for predicting RV and HR simultaneously. Given that fluctuations in respiration and heart rate evoke delayed responses in the BOLD signal (i.e., the BOLD effects occur asynchronously and persist beyond the duration of the physiological response) (Birn et al, 2008; Chang et al, 2009), a neural network that takes past and future information into account – such as bi-LSTM – means that no hemodynamic lag needs to be explicitly considered. Further, the ability to integrate over multiple time-scales may offer benefits over fixed-windowed methods (Salas et al, 2021).

Given that RV and HR are strongly coupled, we had hypothesized that learning the features of one (i.e. RV) could boost the accuracy of predicting the other (i.e. HR) even with a few learning instances. However, experiments with pre-training a network based on RV and fine-tuning these weights to predict HR (compared to training with random weights), see Supplementary Section: Handling Missing Data During Training, resulted in only a small performance increase, and further experiments may be warranted. Commonly, transfer learning is performed by freezing the early layers (high-level features) and training the remaining layers (low-level features) to specialize on another task. Since our network architecture comprises only two levels, a bi-LSTM layer and linear layers, fine-tuning only the LSTM layer may not be enough to transfer the knowledge between RV and HR and needs further investigation with different architectures. This direction is of particular interest given that the measured HR signal is often noisy.

Here, we adopted a large dataset to train and evaluate our models. Although an initial quality-check was performed on physiological recordings, visual inspection indicated that many still contained artifacts such as clipping and imperfect heart-beat detection. However, substantial improvement in performance beyond recent published work (from medians of approximately $r \sim 0.5$ (Bayrak et al, 2020; Salas et al, 2021) to $r \sim 0.7$ for RV in the current study) was attained. (Of note, fMRI data has low signal-to-noise ratio. Given that meaningful fluctuations comprise about 1-4% in the data, correlations at and above 0.5-0.6 between fMRI and an external measure are considered to be relatively strong in the neuroscience field.) In the previous studies, we used a small, visually inspected subset of the HCP data consisting only of RV signals. The aforementioned improvement suggests that a larger, albeit less strictly vetted, dataset was effective for learning physiological patterns. Nonetheless, while a wide variety of signals (including some low-quality examples) may be advantageous during training, low-quality physiological data colors measurements of performance during testing, as these physiological recordings are taken as ground truth. In other words, low correlation scores can emerge when there are artifacts in the physiological recordings, necessitating close inspection as in Figure 8.

White matter regions in the brainstem were shown to be very predictive of breathing as well as heart rate. While much remains to explore in terms of regional

predictiveness of physiological signals beyond the current heuristic approach, the strong performance of tractography-based ROIs suggest that incorporating parcellation boundaries derived from other imaging modalities (e.g., diffusion-weighted MRI) and non-gray-matter regions could benefit fMRI prediction tasks. Further, results of the percent ROI models (Section 2.5) demonstrate that performance could be boosted by increasing the number of ROIs that are included in the training. While this finding supports the need for comprehensive atlases with finer subdivisions, the exponentially decaying projection in Figure 5b suggests that dimensionality reduction using atlases sufficiently captures low-frequency physiological information.

The present work is directly inspired by reports showing that low-frequency physiological processes can impact fMRI analyses (Xifra-Porxas et al, 2021; Power et al, 2017; Birn et al, 2006; Wise et al, 2004). Consistent with prior work, we found that modeling physiological effects could alter maps of functional connectivity, and that RV and HR signals can account for substantial temporal variation in certain brain regions. Previous studies have also found that in task conditions, changes in cerebral oxygenation, captured in part by respiration and heart rate signals, may modulate the magnitude of observed task activation responses (Bright et al, 2020). These findings suggest that accounting for RV and HR signals can advance individual-level precision in neuroscience and medicine using fMRI. As such, it is important to have general, versatile models that can reconstruct these features in the common scenario of missing or noisy physiological recordings.

Notably, in addition to its use for 'denoising' - i.e., removing RV and HR effects from fMRI signals, the proposed method may also open possibilities for studying *neural* processes relating to autonomic regulation in data where physiological signals have not been monitored. Specifically, we may expect that a component of the reconstructed RV and HR signals may track large-scale neural activity that is linked with physiological modulations (e.g., (Tu and Zhang, 2022)). Based on our experiments with individual ROI models, we observed that cortical regions with the highest individual predictive power tended to span somatomotor, salience, and ventral attention regions, which have known connections with autonomic regulation (Mckay et al, 2003; Beissner et al, 2013). While these autonomic signals may not be of interest for some studies, they may inform work on brain-body interactions.

The current results suggest a number of future directions. One springs from limitations in the use of population-level atlases for defining fMRI parcellations. The present study used atlases that had been derived from populations, yet individuals can exhibit variability in the spatial boundaries of functional regions (Kong et al, 2021). Thus, individual- or cohort-specific parcellations may enable higher precision in extracting personalized physiological information from fMRI. The issue of precision is amplified when working with datasets from different stages of human life, including infant or aging datasets, which do not conform with the topographical features of the young cohort that is used to train networks in this work. Another limitation that can be addressed in future work is model interpretability, a common problem with machine learning models. Methods such as Shapley Values (Ghorbani and Zou, 2019) can offer explanations for model predictions and provide insights into

the model development process. Architectural modularity (multiple layers) could be considered to allow transfer learning (Chollet, 2021), since the ability to freeze layers that capture high-level information may further boost the performance of joint learning.

4 Conclusions

Modeling physiological variability is increasingly recognized as crucial for both (1) improving the sensitivity of fMRI to neural effects, as well as (2) providing valuable information about cerebrovascular health, brain states, and emotion regulation. Since high-quality physiological signals are often missing from fMRI datasets, the present study fills this gap by introducing a generalizable tool for reconstructing two key physiological signals (RV, HR) from fMRI data. The proposed framework was found to be robust across a broad range of experimental conditions and imaging protocols, indicating that it can enrich a broad array of fMRI datasets with missing physiological information. This study motivates future work on methodological advances for modeling RV/HR, and enables retrospective studies of physiological effects in health and disease, leveraging the large body of fMRI databases that have been acquired without physiological signals.

5 Methods

5.1 Datasets

Models were trained on the Human Connectome Project (HCP) data (Van Essen et al, 2013). Our models can be divided into two categories, resting state models (rs-models) and task models (t-model). The rs-models were trained using HCP resting state data using 5-fold cross-validation. To evaluate the generalizability of rs-models across subjects both within (resting state) and between (task vs. resting state) conditions, rs-models were tested on both HCP task data as well as a separate, in-house dataset with different acquisition parameters. Similarly, the t-model was trained on HCP task datasets, and tested on held-out HCP task data (with different task conditions), HCP resting state data, and the in-house dataset. The individual datasets and acquisition parameters are described below. For more details, please refer to the published articles (WU-Minn, 2017; Goodale et al, 2021).

5.1.1 Dataset 1

A set of resting-state fMRI (rs-fMRI) scans was drawn from the publicly available HCP 1200 subject release. fMRI scans in this release were acquired using a simultaneous multi-slice EPI sequence with the following parameters: TR = 0.72 s, duration of 14.33 mins, voxel size of 2 mm isotropic, TE = 33.1 ms, multi-band factor = 8, flip angle = 52 deg and 72 slices. During the resting-state scans, subjects were instructed to keep their eyes open and fixate on a cross-hair, and subjects underwent 4 scans (two runs on one day, and two runs on a second day). A subset of the HCP rs-fMRI dataset, consisting only of those subjects whose physiological signals were reported

to have passed a quality assessment in both Power et al. (Power et al, 2020) and Xifra-Porxas et al. (Xifra-Porxas et al, 2021), was included in this study. This procedure resulted in N = 375 subjects (with all 4 runs, totalling 1500 scans).

5.1.2 Dataset 2

Our study also drew upon task fMRI (t-fMRI) scans from the HCP 1200 subject release. These scans were acquired with the same parameters as the rs-fMRI data, described above, except for the duration of the scans. The task dataset comprises seven tasks: Working Memory, Social Cognition, Emotion Processing, Gambling, Relational Processing, Motor, Language. These tasks were designed to activate a variety of brain networks, and each subject underwent 2 scans per task (both runs on the same day). signals was one that had excluded that subject from the All scans from t-fMRI datasets were initially considered. A quality assessment was employed, using automated selection criteria to remove scans with poor-quality physiological data (this script will be found here upon acceptance: http://github.com/deep-physio-recon/qa). Briefly, these criteria checked for clipping of waveforms (values clamped at 0 and 4095), unrealistic heart rates (mean heart rates below 30 bpm or above 97 bpm, or constant at 48 bpm (Xifra-Porxas et al, 2021), and missing waveforms. A subset of fMRI scans for each task, consisting of only those scans whose physiological data were labeled as 'clean' by the above criteria, was included in this study. For detailed information about task fMRI scans, please see Supplementary Table 2.

5.1.3 Dataset 3

A different, in-house dataset was used as an external validation dataset. fMRI data were acquired with a multi-echo, gradient-echo EPI sequence, with the following parameters: TR = 2.1 s, duration of 24.5 mins, voxel size of 3 mm isotropic, echo times of [13.0, 29.4, 47.5 ms], flip angle = 75 deg, 30 axial slices (for detailed acquisition information see (Goodale et al, 2021)). The scans were acquired under 5 different conditions: cued deep breaths, eyes-open auditory task, eyes-closed auditory task, passive eyes-closed auditory task (i.e. no button press) and eyes-closed rest (no stimuli presented). In the auditory tasks, auditory tones were delivered at long inter-stimulus intervals, and subjects were instructed to press a button as quickly as possible. We selected a random subset of these scans that additionally underwent visual inspection to have clean physiological data. This yielded a total of 23 scans (drawn from 11 subjects).

5.2 Preprocessing

5.2.1 fMRI

Both resting-state and task fMRI scans had undergone the HCP minimal preprocessing pipeline (Glasser et al, 2013). Beyond this, we applied linear and quadratic detrending to remove slow scanner drifts. This was followed by band-pass filtering, within the low-frequency range of (0.01 - 0.15 Hz) and temporal downsampling by a factor of 2 (which does not result in further information loss, since the Nyquist

criterion is satisfied). The HCP minimal preprocessing pipeline included motion coregistration, and while an additional step of regressing out head motion parameters is an integral part of many fMRI preprocessing pipelines, we did not regress out head motion parameters when training our models since they may carry predictive information about physiological fluctuations. For example, recent evidence has shown that apparent head motions can be introduced by respiration (Power et al, 2019), and we have previously found that retaining head motion slightly improved the reconstruction of respiration variation signal (Salas et al, 2021). Therefore, head motion was retained in the training data, to preserve information relevant to predicting RV and HR. The in-house data were preprocessed according to a multi-echo ICA pipeline described in (Goodale et al, 2021). All fMRI scans were already spatially aligned to the common MNI152 space.

5.2.2 Physiological Recordings

From the pulse oximetry signal, heart rate (HR) was extracted as the inverse of the mean inter-beat-interval in sliding windows of 6 seconds centered at each fMRI time frame (TR). Likewise, the respiration variation (RV) signal was calculated as the temporal standard deviation of the raw respiration waveform in a window of 6 sec centered at each TR (Chang et al, 2009). Both RV and HR were then band-pass filtered (0.01 – 0.15 Hz) and were resampled to temporal resolution of 1.44 seconds, in a manner identical to the HCP fMRI data, the fMRI data that is used to train the models.

5.2.3 Normalization

fMRI signals and respiration belt data were both acquired with arbitrary units and may carry scan- and subject-specific amplitude differences. Unless stated otherwise, all time-series signals, including heart rate, were temporally normalized to zero mean and unit variance.

5.2.4 Dimensionality Reduction

When training a neural network on voxelwise, 4D whole-brain data, downsampling or patch/window-based implementations are typically required due to GPU memory limitations. Extracting time courses that are averaged within functional or anatomically defined regions of interest (ROIs) enables computationally efficient modeling. The use of ROI time courses also renders the approach less sensitive to the spatial resolution of the acquired fMRI data. Therefore, here we carried out dimensionality reduction by parcellating the brain into regions of interest based on four published atlases. Of note, our learning framework can flexibly accommodate any set of atlas regions for dimensionality reduction. The four published atlases used here were derived from multiple imaging modalities: a cerebral cortex atlas that was derived from rs-fMRI data provided 400 cortical regions embedded within 7 larger functional networks (Schaefer et al, 2018); the Pandora TractSeg white matter atlas, which was derived from diffusion MRI data and included 72 uni-/bi-lateral white matter regions (Hansen et al, 2020); the Melbourne subcortex atlas, which provides a multi-modal

segregation of 16 subcortical regions (Tian et al, 2020); and an ascending arousal network (AAN) atlas that includes 9 regions located in the brainstem (Edlow et al, 2012). Dimensionality and noise were reduced by extracting the mean fMRI time series from all voxels within each ROI.

All four atlases were already registered to MNI152 space, and when necessary, were also resampled to 2 mm isotropic voxels to match the resolution of the preprocessed fMRI scans. While the Pandora white matter atlases (Hansen et al, 2020) includes various options, a probabilistic atlas created using TractSeg method (Wasserthal et al, 2018) on the HCP cohort was utilized in this study. We thresholded the probabilistic atlas (at 95%) to exclude voxels with lower confidence and minimize the overlap between white matter and gray matter regions. The 7 Tesla HCP Scale I atlas from Melbourne subcortex atlases (Tian et al, 2020) was selected, with the assumption that 7T may provide better spatial precision and the Scale I granularity could support generalizability to external cohorts.

5.3 Network Architecture and Implementation Details

Deep neural networks have shown remarkable success for image and time-series data (Sharma and Singh, 2017), including in the field of fMRI. The success of these networks stems from their ability to find a non-linear representation of the data and to make meaningful connections between spatial and temporal information implicit in the data. The proposed framework consists of a bidirectional LSTM (bi-LSTM) network followed by dropout and two linear layers. Long short term memory networks (LSTM) were selected as a candidate approach based on their capacity to automatically learn the temporal dependencies present in time series and their capability of operating on data of varying lengths (Van Houdt et al, 2020). Linear layers are commonly used for inferring an objective-specific feature space, and here they were used to infer unique RV and HR estimates from LSTM hidden units. The framework is illustrated in Figure 1. The input to the networks consist of ROI time-series signals provided as different channels (# of channels = # of ROIs). The bi-LSTM network is followed by a dropout layer and is connected to two linear layers, which output the estimated RV and HR time series. These estimated signals are the same length as the corresponding input ROI time series.

We conducted a grid search to select hyperparameters for the resting-state models. The size of hidden states was selected from values of [32, 64, 512, 1000, 1024, 2000, 2048], batch size from values of [1, 2, 8, 16, 32, 64], for dropout rates of [0.1, 0.3, 0.5, 0.6], for learning rates from [1.0e-2, 1.0e-3, 1.0e-4, 1.0e-5] and for the decaying learning rate with decay rate of [0.01, 0.1, 0.2, 0.5].

All models were then trained with the following, empirically chosen hyperparameters: a hidden state (h) size 2000 (i.e., depth), a batch size of 16, dropout rate 0.3 and were trained with decaying learning rate (lr) of 1.0e-3 with patience 2 and decay rate of 0.5, saving only the best models according to validation performance. The models were trained using ADAM optimizer with default parameters. The experiments were performed on an NVIDIA RTX 2080Ti GPU. Programs were implemented with Python using the Pytorch deep learning library. The code for the model training

framework will be publicly available at https://github.com/neurdylab/deep-physio-recon/journal-name.

5.4 Model training and evaluation

Resting state (rs-) models. Only for the model training on the HCP resting-state dataset (Datasets: Dataset 1), we used 5-fold cross validation, with 68% for training and 12% for validation and 20% for testing. A subset of 375 subjects (1500 scans) were used for training by rotating the partitions (Supplementary Figure 5; left), and the resulting performance (pooled over the 5 testing partitions) is reported. In addition, to evaluate the generalizability across subjects both within (resting state) and between (task vs. resting state) conditions, rs-models were tested on HCP task (Dataset 2) and in-house (Dataset 3) datasets. Since the 5-fold CV on the rs-fMRI data resulted in 5 different models, one of these models was randomly selected to be applied to the task and in-house datasets (Supplementary Figure 5; right). For detailed information about number of scans used for training and testing, please see Supplementary Table 3.

Training task (t-fMRI) model. t-fMRI model is trained using HCP task fMRI dataset (Dataset 2). More specifically, models were trained using 4 tasks (working memory, social cognition, emotional processing, and gambling tasks), and assessed using the remaining 3 tasks (relational processing, motor and language tasks). Of note, the subjects that were included in the training t-fMRI model were excluded from testing cohorts (Supplementary Figure 6; left), in order to assess generalizability across different subjects. In addition, to further evaluate the generalizability across subjects and between (task vs. resting state) conditions, t-models were also tested on HCP resting (Dataset 1) and in-house (Dataset 3) datasets (Supplementary Figure 6; right). For detailed information about number of scans used for training and testing, please see Supplementary Table 4.

5.5 Impact on fMRI signals and network connectivity

5.5.1 Percent Variance

To assess the degree to which our reconstructed RV and HR signals could account for fMRI signal fluctuations across the brain, we examined the percentage of temporal variance explained in each fMRI voxel signal by the reconstructed RV and HR signals. The predicted time courses were first convolved with a previously determined transfer function (for RV, respiration response function (Birn et al, 2008); for HR, cardiac response function (Chang et al, 2009)) that captures the forward mapping between physio and fMRI fluctuations, as well as their time and dispersion derivatives (Chen et al, 2020) to allow for small deviations in latency and shape from the canonical model (Henson et al, 2002). The percent variance explained was defined as the fraction by which a voxel's original temporal variance would be reduced after projecting out (via ordinary least squares) a linear combination of the aforementioned regressors, and multiplying by 100. For each scan, the RV and HR predictions used in this analysis were obtained from cross-validated rs-models, where the model that was applied to derive a subject's predicted RV/HR signals was one that had excluded

that subject from the training. This analysis used the HCP resting-state data that has been bandpass filtered and downsampled by a factor of 2 (see Section 5.2). The percent variance maps calculated from individual scans were then averaged to get the population-level mapping of the RV and HR effects.

5.5.2 Connectivity Analysis

For both the ROI-based and seed-based connectivity analyses, we used resting-state scans that have undergone the HCP ICA-FIX preprocessing pipeline (Griffanti et al, 2014). The rationale behind using ICA-FIX data for this analysis is to show the utility of physiological signals even beyond ICA-FIX denoising, as substantial RV and HR effects may remain. Beyond ICA-FIX, we applied linear and quadratic detrending to remove slow scanner drifts. This was followed by band-pass filtering within the low-frequency range of (0.01 - 0.15 Hz) and temporal downsampling by a factor of 2.

ROI-based. To assess the impact of the physiological signal of interest on regional correlations, we pursued a functional connectivity analysis. Functional connectivity commonly refers to similarities in brain activity signals between regions, and is calculated as the pairwise (Pearson) correlation of ROI time series and represented as a symmetrical matrix. We assessed the pairwise functional connectivity between all 497 ROIs drawn from cortical, white matter, subcortical, and ascending arousal network atlases (Section 5.2.4) both before and after projecting out the estimated physiological subspace using the same basis functions described above (see Section 5.5.1), as well as 6 rigid-body head motion parameters and their derivatives. The Pearson correlation between each pair of ROIs were calculated and used to construct 497 x 497 symmetrical connectivity matrices where each element represents temporal similarity score between two ROIs.

Seed-based. Seed-based analysis is one of the most common ways to explore functional connectivity within the brain (Van Den Heuvel and Pol, 2010). In a seed-based correlation analysis, connectivity is calculated as the correlation between the time course of a selected reference ("seed") region to all other voxels in the brain. The resulting connectivity map represents the Pearson correlation scores for each voxel, indicating how well each voxel's time series correlates with the time series of the seed. In this analysis, we pursued seed-based connectivity with respect to a seed region in the default mode network (DMN). For comparison, these analyses were repeated after projecting out the measured and predicted RV and HR waveforms, using the basis sets and derivatives described above (see Section 5.5.1), as well as 6 rigid-body head motion parameters and their derivatives.

Supplementary information. The article has accompanying supplementary files.

Declarations

Funding.

This work was supported by NIH grants K22ES028048 and RF1MH125931.

Conflict of interest/Competing interests.

No competing interests to declare.

Availability of data and materials.

Dataset 1 and Dataset 2 used in the present study are available for download from the Human Connectome Project (www.humanconnectome.org). Dataset 3 is available for download from OSF. Moreover, pretrained model weights and predicted RV and HR signals will be publicly available upon publication (detailed in the online documentation for more details) at github.com/neurdylab/deep-physio-recon/journal-name.

Code availability.

DeepPhysioRecon model code will be found on GitHub upon acceptance (github.com/neurdylab/deep-physio-recon/journal-name), along with user documentation. Code for data pre-/post-processing and analysis will be shared as Colab notebooks that illustrate the use of DeepPhysioRecon using exemplary data.

References

Ash T, Suckling J, Walter M, et al (2013) Detection of physiological noise in resting state fmri using machine learning. Human brain mapping 34(4):985–998 ↪ page 3

Aslan S, Hocke L, Schwarz N, et al (2019) Extraction of the cardiac waveform from simultaneous multislice fmri data using slice sorted averaging and a deep learning reconstruction filter. Neuroimage 198:303–316 ↪ page 3

Azzalini D, Rebollo I, Tallon-Baudry C (2019) Visceral signals shape brain dynamics and cognition. Trends in cognitive sciences 23(6):488–509 ↪ page 2

Bandettini PA, Wong EC, Hinks RS, et al (1992) Time course epi of human brain function during task activation. Magnetic resonance in medicine 25(2):390–397 ↪ page 2

Barrett LF, Simmons WK (2015) Interoceptive predictions in the brain. Nature reviews neuroscience 16(7):419–429 ↪ page 2

Bayrak RG, Salas JA, Huo Y, et al (2020) A deep pattern recognition approach for inferring respiratory volume fluctuations from fmri data. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, pp 428–436 ↪ pages 3 and 15

Bayrak RG, Hansen CB, Salas JA, et al (2021) From brain to body: Learning low-frequency respiration and cardiac signals from fmri dynamics. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, pp 553–563 ↪ page 3

Beall EB, Lowe MJ (2007) Isolating physiologic noise sources with independently determined spatial measures. Neuroimage 37(4):1286–1300 ↪ page 3

Beissner F, Meissner K, Bär KJ, et al (2013) The autonomic brain: an activation likelihood estimation meta-analysis for central processing of autonomic function. Journal of neuroscience 33(25):10,503–10,511 ↪ page 16

Bernier M, Cunnane SC, Whittingstall K (2018) The morphology of the human cerebrovascular system. Human brain mapping 39(12):4962–4975 ↪ page 9

Birn RM, Diamond JB, Smith MA, et al (2006) Separating respiratory-variation-related fluctuations from neuronal-activity-related fluctuations in fmri. Neuroimage 31(4):1536–1548 ↪ pages 2, 12, and 16

Birn RM, Smith MA, Jones TB, et al (2008) The respiration response function: the temporal dynamics of fmri signal fluctuations related to changes in respiration. Neuroimage 40(2):644–654 ↪ pages 15 and 21

Birn RM, Cornejo MD, Molloy EK, et al (2014) The influence of physiological noise correction on test–retest reliability of resting-state functional connectivity. Brain connectivity 4(7):511–522 ↪ page 14

Botvinik-Nezer R, Holzmeister F, Camerer CF, et al (2020) Variability in the analysis of a single neuroimaging dataset by many teams. Nature 582(7810):84–88 ↪ page 2

Bright MG, Whittaker JR, Driver ID, et al (2020) Vascular physiology drives functional brain networks. NeuroImage p 116907 ↪ pages 2 and 16

Bycroft C, Freeman C, Petkova D, et al (2018) The uk biobank resource with deep phenotyping and genomic data. Nature 562(7726):203–209 ↪ page 3

Chai XJ, Castañón AN, Öngür D, et al (2012) Anticorrelations in resting state networks without global signal regression. Neuroimage 59(2):1420–1428 ↪ page 6

Chang C, Glover GH (2009) Effects of model-based physiological noise correction on default mode network anti-correlations and correlations. Neuroimage 47(4):1448–1459 ↪ page 6

Chang C, Cunningham JP, Glover GH (2009) Influence of heart rate on the bold signal: the cardiac response function. Neuroimage 44(3):857–869 ↪ pages 2, 12, 15, 19, and 21

Chen JE, Lewis LD, Chang C, et al (2020) Resting-state "physiological networks". NeuroImage 213:116,707 ↪ pages 2 and 21

Chollet F (2021) Deep learning with Python. Simon and Schuster ↪ page 17

Dubois J, Adolphs R (2016) Building a science of individual differences from fmri. Trends in cognitive sciences 20(6):425–443 ↪ page 2

Edlow BL, Takahashi E, Wu O, et al (2012) Neuroanatomic connectivity of the human ascending arousal system critical to consciousness and its disorders. Journal of Neuropathology & Experimental Neurology 71(6):531–546 ↪ page 20

Ghorbani A, Zou J (2019) Data shapley: Equitable valuation of data for machine learning. In: International Conference on Machine Learning, PMLR, pp 2242–2251 ↪ page 16

Glasser MF, Sotiropoulos SN, Wilson JA, et al (2013) The minimal preprocessing pipelines for the human connectome project. Neuroimage 80:105–124 ↪ page 18

Glasser MF, Coalson TS, Bijsterbosch JD, et al (2018) Using temporal ica to selectively remove global noise while preserving global signal in functional mri data. NeuroImage 181:692–717 ↪ page 3

Goodale SE, Ahmed N, Zhao C, et al (2021) fmri-based detection of alertness predicts behavioral response variability. Elife 10:e62,376 ↪ pages 17, 18, and 19

Griffanti L, Salimi-Khorshidi G, Beckmann CF, et al (2014) Ica-based artefact removal and accelerated fmri acquisition for improved resting state network imaging. Neuroimage 95:232–247 ↪ page 22

Hansen CB, Yang Q, Lyu I, et al (2020) Pandora: 4-d white matter bundle population-based atlases derived from diffusion mri fiber tractography. bioRxiv ↪ pages 19 and 20

Henson RN, Price CJ, Rugg MD, et al (2002) Detecting latency differences in event-related bold responses: application to words versus nonwords and initial versus repeated face presentations. Neuroimage 15(1):83–97 ↪ page 21

Jack Jr CR, Barnes J, Bernstein MA, et al (2015) Magnetic resonance imaging in alzheimer's disease neuroimaging initiative 2. Alzheimer's & Dementia 11(7):740–756 ↪ page 3

Kassinopoulos M, Mitsis GD (2019) Identification of physiological response functions to correct for fluctuations in resting-state fmri related to heart rate and respiration. Neuroimage 202:116,150 ↪ page 12

Koban L, Gianaros PJ, Kober H, et al (2021) The self in context: brain systems linking mental and physical health. Nature Reviews Neuroscience 22(5):309–322 ↪ page 2

Kong R, Yang Q, Gordon E, et al (2021) Individual-specific areal-level parcellations improve functional connectivity prediction of behavior. Cerebral Cortex

31(10):4477–4500 ↪ page 16

Kwong KK, Belliveau JW, Chesler DA, et al (1992) Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. Proceedings of the National Academy of Sciences 89(12):5675–5679 ↪ page 2

Mckay LC, Evans KC, Frackowiak RS, et al (2003) Neural correlates of voluntary breathing in humans. Journal of applied physiology 95(3):1170–1178 ↪ page 16

Murphy K, Birn RM, Bandettini PA (2013) Resting-state fmri confounds and cleanup. Neuroimage 80:349–359 ↪ page 2

Nashiro K, Min J, Yoo HJ, et al (2022) Increasing coordination and responsivity of emotion-related brain regions with a heart rate variability biofeedback randomized trial. Cognitive, Affective, & Behavioral Neuroscience pp 1–18 ↪ page 2

Noble S, Scheinost D, Constable RT (2019) A decade of test-retest reliability of functional connectivity: A systematic review and meta-analysis. Neuroimage 203:116,157 ↪ page 6

Ogawa S, Tank DW, Menon R, et al (1992) Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. Proceedings of the National Academy of Sciences 89(13):5951–5955 ↪ page 2

Picchioni D, Özbay PS, Mandelkow H, et al (2022) Autonomic arousals contribute to brain fluid pulsations during sleep. NeuroImage p 118888 ↪ page 2

Power JD, Plitt M, Laumann TO, et al (2017) Sources and implications of whole-brain fmri signals in humans. Neuroimage 146:609–625 ↪ pages 2, 12, and 16

Power JD, Lynch CJ, Silver BM, et al (2019) Distinctions among real and apparent respiratory motions in human fmri data. NeuroImage 201:116,041 ↪ page 19

Power JD, Lynch CJ, Dubin MJ, et al (2020) Characteristics of respiratory measures in young adults scanned at rest, including systematic changes and "missed" deep breaths. Neuroimage 204:116,234 ↪ page 18

Salas JA, Bayrak RG, Huo Y, et al (2021) Reconstruction of respiratory variation signals from fmri data. NeuroImage 225:117,459 ↪ pages 3, 15, and 19

Schaefer A, Kong R, Gordon EM, et al (2018) Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri. Cerebral cortex 28(9):3095–3114 ↪ page 19

Sharma P, Singh A (2017) Era of deep neural networks: A review. In: 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), IEEE, pp 1–5 ↪ page 20

Shmueli K, van Gelderen P, de Zwart JA, et al (2007) Low-frequency fluctuations in the cardiac rate as a source of variance in the resting-state fmri bold signal. Neuroimage 38(2):306–320 ↪ page 2

Shokri-Kojori E, Tomasi D, Volkow ND (2018) An autonomic network: synchrony between slow rhythms of pulse and brain resting state is associated with personality and emotions. Cerebral Cortex 28(9):3356–3371 ↪ page 2

Tian Y, Margulies DS, Breakspear M, et al (2020) Hierarchical organization of the human subcortex unveiled with functional connectivity gradients. bioRxiv ↪ page 20

Tong Y, Hocke LM, Frederick BB (2019) Low frequency systemic hemodynamic "noise" in resting state bold fmri: characteristics, causes, implications, mitigation strategies, and applications. Frontiers in neuroscience p 787 ↪ page 2

Tu W, Zhang N (2022) Neural underpinning of a respiration-associated resting-state fmri network. ELife 11:e81,555 ↪ page 16

Van Den Heuvel MP, Pol HEH (2010) Exploring the brain network: a review on resting-state fmri functional connectivity. European neuropsychopharmacology 20(8):519–534 ↪ pages 6 and 22

Van Essen DC, Smith SM, Barch DM, et al (2013) The wu-minn human connectome project: an overview. Neuroimage 80:62–79 ↪ page 17

Van Houdt G, Mosquera C, Nápoles G (2020) A review on the long short-term memory model. Artificial Intelligence Review 53(8):5929–5955 ↪ page 20

Wasserthal J, Neher P, Maier-Hein KH (2018) Tractseg-fast and accurate white matter tract segmentation. NeuroImage 183:239–253 ↪ page 20

Wise RG, Ide K, Poulin MJ, et al (2004) Resting fluctuations in arterial carbon dioxide induce significant low frequency variations in bold signal. Neuroimage 21(4):1652–1664 ↪ pages 2 and 16

WU-Minn H (2017) 1200 subjects data release reference manual. URL https://www humanconnectome org ↪ pages 3 and 17

Xifra-Porxas A, Kassinopoulos M, Mitsis GD (2021) Physiological and motion signatures in static and time-varying functional connectivity and their subject identifiability. Elife 10:e62,324 ↪ pages 2, 16, and 18

Zhang J, Kucyi A, Raya J, et al (2021) What have we really learned from functional connectivity in clinical populations? NeuroImage 242:118,466 ↪ page 6