Predictive Modeling of Brain-Body Association

Minkyu Choi *
EECS, Univeristy of Michigan
cminkyu@umich.edu

Xiaokai Wang *
BME, University of Michigan
xiaokaiw@umich.edu

Zhongming Liu
EECS, BME, University of Michigan
zmliu@umich.edu

Abstract

The brain senses and regulates internal organs through a process called interoception. Appealing theoretical frameworks have been proposed to describe this dynamic interaction, yet they are not supported by computational models that can be validated with experimental data. In this paper, we develop a foundation model to substantiate the embodied predictive interoception coding theory. Through self-supervised learning, the model learns the dynamics of the physiological data and captures their association with brain activities. The model reveals a bilateral and distributed brain network across subcortical to cortical levels that encodes respiratory and cardiac dynamics by predicting their states forward in seconds to tens of seconds. As an initial attempt, our work merits future studies in modeling interoception through concurrent brain and body measurements.

1 Introduction

Foundation models trained with large-scale datasets can support diverse downstream tasks, as demonstrated in natural language processing [5, 2], computer vision [7, 20], and multimodal fusion [19, 12]. Similar development has also emerged for biomedical applications [16, 24]. To name a few examples, foundation models trained on physiological signals help detect cardiovascular diseases [17, 18]. Separate models trained on brain functional magnetic resonance imaging (fMRI) data capture brain networks and yield representations associated with behaviors [3, 6, 13]. However, there is no established model that links brain and bodily signals through multimodal learning. This gap is notable because the brain and the body interact to maintain both mental and physical health, a process known as interoception. Disruption of brain-body interactions often leads to chronic conditions that are resistant to treatments targeting either the brain or the body alone. Additionally, methods for training models of brain-body association should utilize learning objectives that are compatible with biological mechanisms or principles, whereas certain engineering designs, such as masking or augmenting input, are undesirable.

We take a whole-person perspective to model brain-body association through predictive learning and modeling. Our model substantiates embodied predictive interoception coding (EPIC), a theory stating that the brain predicts future bodily states through a hierarchy of neural pathways [1]. While this theory has gained popularity in neuroscience, including embodied cognition [22, 21], it remains largely conceptual and lacks learnable computational models. As an initial attempt, our model learns a latent representation that generates cardiac and respiratory signals and predicts their dynamics, and associates the resulting latent representations with fMRI signals across subcortical to cortical regions.

*These authors contribute equally to the work.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: Foundation Models for the Brain and Body.

The model demonstrates a unique capability of mapping and quantifying brain-body association based on simultaneous brain fMRI and physiological recordings.

2 Datasets

We use public datasets from the Human Connectome Project (HCP) [23] to train and evaluate the model. The dataset includes concurrently acquired physiological recordings (cardiac and respiratory signals) and resting-state functional magnetic resonance imaging (rs-fMRI). Physiological recordings are used to train the model for learning meaningful representations and modeling their temporal dynamics, whereas rs-fMRI is used to examine the association between brain activities and bodily representations. After excluding sessions with missing or corrupted physiological recordings, we randomly assign the remaining data into disjoint training, validation, and test sets comprising 658, 46, and 183 subjects, respectively.

rs-fMRI data Data were collected while participants lay in a 3-T magnetic resonance imaging (MRI) scanner with eyes open, fixating on a crosshair, and refraining from performing any explicit task. Each rs-fMRI session lasted approximately 15 minutes. We use the data corrected with minimal preprocessing steps [9], including distortion correction, spatial normalization, cortex surface generation and registration, and temporal resampling at 1 Hz.

Physiological data Cardiac and respiratory signals were recorded simultaneously with rs-fMRI using photoplethysmograph (PPG) and a pneumatic respiration belt, respectively. These signals were originally sampled at 500 Hz and subsequently down-sampled to 32 Hz to reduce computational cost while preserving relevant temporal features.

3 Methods

We build a generative model to learn compact representations of physiological signals and model their temporal dynamics. The model consists of two main components: a variational autoencoder (VAE) [14] for encoding instantaneous bodily states and a temporal encoder based on gated recurrent units (GRU)s [4] for predicting forward in time. This design draws inspiration from the EPIC theory: the hierarchical neural pathways between the brain and internal organs allow the brain to sense, predict, and adjust future physiological states through active inference (or allostasis) [1] (Fig. 1(a)). Training proceeds in two self-supervised stages. The trained model is used to analyze brain and physiological states in the held-out test set, providing a computational tool to map and quantify brain-body association. The following sections describe model architecture (Section 3.1), learning strategies (Section 3.2), and evaluation (Section 3.3).

3.1 Model Architecture

Variational autoencoder The VAE encodes instantaneous bodily states into latent vectors $\mathbf{z}$. Its architecture mirrors the hierarchical neural pathway between the brain and the internal organs, as illustrated in Fig. 1(b). For each organ, an organ-specific encoder transforms a 1-D input signal from a 16-second window (slid forward by 1 second per time) into a feature vector. This encoder uses four convolutional layers, each performing 1-D convolution followed by batch normalization and ReLU activation. The organ-specific representations are then concatenated as the input to a shared organ encoder, which learns the joint representation across organs. This shared encoder uses eight layers of 1-D convolution (with batch normalization and ReLU nonlinearity) and produces a single latent vector $\mathbf{z}$ (256-dimension) for cardiac and respiratory signals within each 16-second window. The decoder transforms the latents back into signals, using an architecture that is symmetric to the encoder. A shared organ decoder (eight layers) maps $\mathbf{z}$ back into concatenated features, which are then passed to organ-specific decoders (four layers each) to reconstruct the original physiological signals.

Temporal encoder The temporal encoder models the dynamics of the latent representations $\mathbf{z}$ from VAE into context vectors $\mathbf{c}$ (256-dimension). It consists of two GRU layers (hidden layer: 128-dimension) followed by a fully connected layer, as in Fig. 1(c). Given a sequence of latent

vectors $\mathbf{z}$ before and at t, the GRU updates its internal state to form a context vector $\mathbf{c}$, which is further mapped by the fully connected layer to predict the immediately next latent vector $\mathbf{z}$ at $t+1$ second. Consecutive latent vectors $\mathbf{z}$ are obtained by sliding the input window of the VAE forward by 1 second.

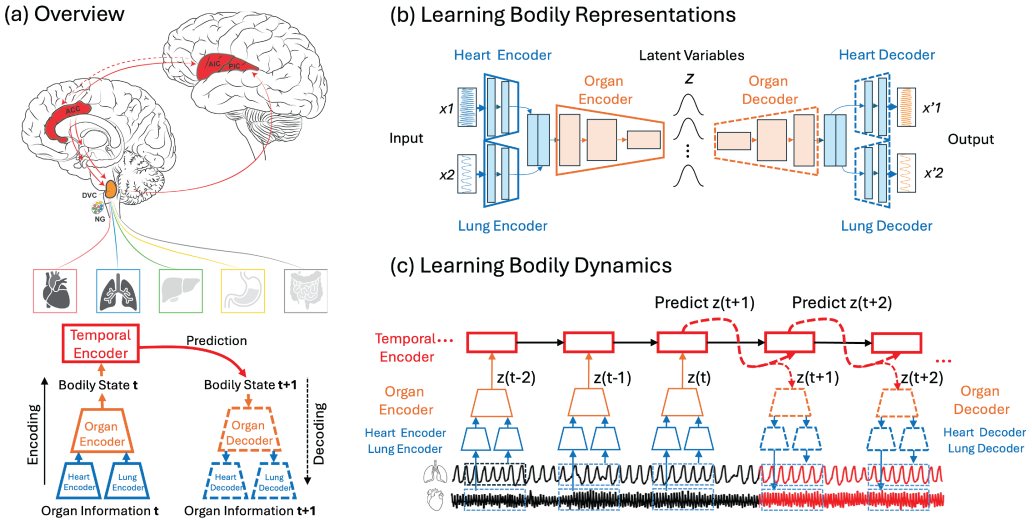

Figure 1: Model structure. (a) Overview of the model architecture, including a variational autoencoder in (b) and a temporal encoder in (c). The correspondence between the brain and the model is shown using matched colors: cortical regions, e.g., AIC/ACC/PIC, are modeled by the temporal encoder (red); the subcortical structures, e.g., the DVC, are modeled by the organ encoder (orange) in the proposed model. For more details on the definitions of the brain regions and their roles in interoception, see Appendix.

3.2 Learning Strategies

Training is done in two stages: first for the VAE without the temporal encoder; second for the temporal encoder with the VAE frozen.

Stage 1 The VAE is trained to (1) faithfully reconstruct the input signals by minimizing the mean squared error (MSE) between the input and the reconstruction, and (2) constrain the latent variables to follow a multivariate standard Gaussian distribution by minimizing the Kullback-Leibler (KL) divergence. The total loss combines MSE and KL divergence terms weighted by a hyperparameter $\beta = 1$ [11]. The VAE model is trained for 100 epochs with the Adam optimizer, with a batch size of 512 and an initial learning rate of 0.01. The learning rate is reduced to 0.001 after the first 50 epochs. After the first 70 epochs, the learning rate is further reduced to 0.0001. Input signals are standardized per subject before training.

Stage 2 With the VAE weights frozen, the temporal encoder is trained to predict the next latent vector $\mathbf{z}$ by minimizing MSE. The model is trained for a total of 60 epochs with the Adam optimizer and a batch size of 512. The initial learning rate is set to 0.001 for the first 40 epochs, reduced to 0.0001 after the 40 epochs, and 0.00001 after 50 epochs.

3.3 Brain-Body Association

We evaluate the brain-body co-fluctuations using canonical correlation analysis (CCA) with the held-out test set. From the temporal encoder, we extract the context vector $\mathbf{c}$ for each time point. The rs-fMRI time series are shifted to sync with the physiological states by accounting for the hemodynamic delay. We compute the covariance matrix between $\mathbf{c}$ ($T \times 256$) and the voxel-wise rs-fMRI ($T \times V$) signals for each subject, where T is the total time length and V is the number of voxels (or vertices) in rs-fMRI. The resulting covariance matrices ($256 \times V$) are averaged across subjects, followed by

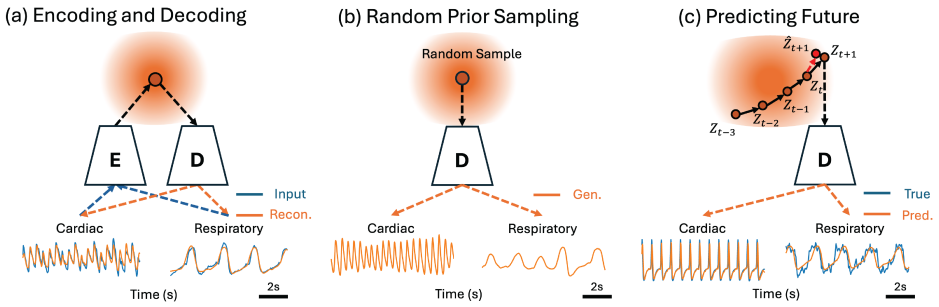

Figure 2: Model validation. The training of the proposed model is validated by evaluating the consistency between input and reconstructed output in (a), the generated physiological signals by randomly sampling from the prior distribution in the latent space in (b), and the temporally predicted physiological signals in (c).

singular value decomposition (SVD) to obtain singular vectors ($COV = U\Sigma V$, where U is for brain, V is for body). The estimated singular vectors are ordered by the corresponding singular values in Σ. The first four singular vectors are retained, capturing the maximal shared variance between brain and bodily signals. Using these components, we reconstruct (or predict) the portion of the fMRI signals that can be explained by physiological states and compute voxel-wise Pearson correlations between the reconstructed and actual fMRI signals. High correlation values indicate brain regions whose activities co-vary with fluctuations in bodily state. A detailed description of the mathematical formulation and statistical tests is provided in the Appendix.

4 Reults

4.1 Model Validation

The performance of the model in reconstructing, generating, and predicting physiological signals is evaluated (Fig. 2). First, our VAE model successfully reconstructs high-fidelity physiological signals. The reconstructed PPG and respiratory traces closely match the measured signals in their shapes (Fig. 2(a)). Specifically, the reconstructed waveform preserved the timing and relative amplitude. Second, the generated PPG and respiratory traces are physiologically plausible when using a latent vector sampled from the prior standard Gaussian distribution. As shown in Fig. 2(b), the synthetic traces closely approximate the empirical data. Lastly, the model accurately predicts future physiological signals from their past history. Fig. 2(c) shows a representative example with a predicted time window of nine seconds. The predicted waveforms retain the correct oscillatory pattern for both the cardiac and respiratory signals, with minimal phase shift. The reconstructive, generative, and predictive abilities of the model ensure that it adequately captures bodily states, as well as their dynamic changes, to study brain-body associations or perform other downstream tasks.

4.2 Brain-Body Association And Interoceptive Network

To systematically quantify the relationship between brain and bodily states, we employ CCA to map associations between the learned physiological representations and voxel-wise brain activity (Section 3.3). We also distinguish brain areas specifically associated with each organ using a selective silencing approach. By inputting zero values for either cardiac or respiratory signals into the VAE encoder, we could separately map brain regions with activities either specifically associated with a single organ or associated with both organs. See Appendix for full details on CCA implementation.

Fig. 3 summarizes the brain-body association uncovered by CCA and selective silencing analyses. The top row visualizes the level of alignment between brain regions and physiological states with correlations. Notably, the integrated network mapped with both modalities (panel a) spans the insular cortex, cingulate cortex, primary somatosensory and motor cortices, and parietal areas. These brain regions have been previously implicated in encoding interoceptive rhythms in both human and animal studies [8]. The network also extends into visual and auditory cortices and substantially overlaps with canonical resting-state networks such as the salience and default mode networks [15].

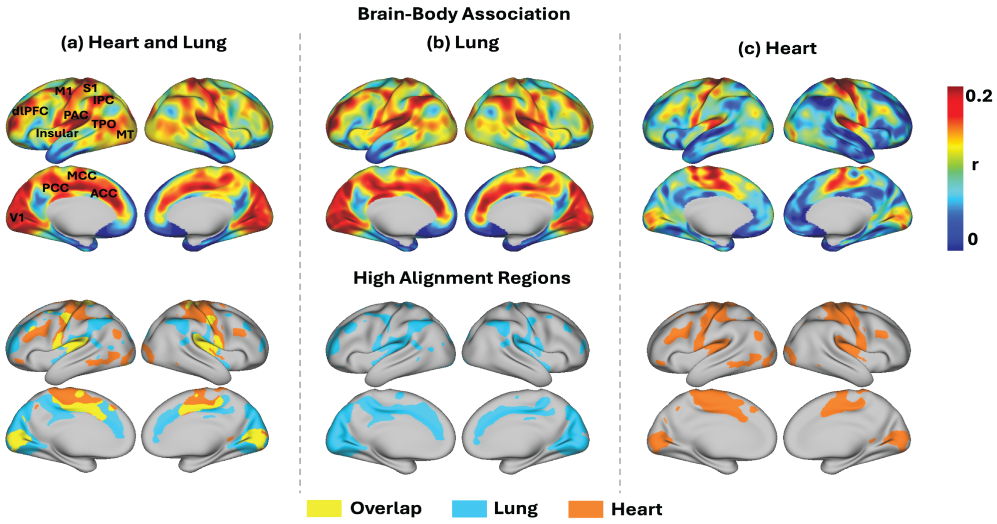

Figure 3: Interoceptive network. The top row illustrates brain regions whose activities closely align with bodily states, reflecting various levels of brain-body association. The brain regions are involved in processing signals from both heart and lung (a), lung alone (b), or heart alone (c). The bottom row shows corresponding brain masks, further highlighting brain regions that show significant association with the bodily states. For more details on the definition of the brain regions, see Appendix.

The brain masks, shown in the bottom row, further highlight the brain regions that are devoted to processing respiratory signals alone (blue, panel b), cardiac signals alone (orange, panel c), or both (yellow, panel a). Interestingly, the primary somatosensory and motor cortices are identified as encoding signals from both modalities (yellow), challenging the classic perspective of a topographical homunculus representation of bodily parts in these regions. See supplementary materials for more results, including subcortical mapping.

5 Discussion

Our work provides a comprehensive mapping of the brain-body association, more extensive than prior studies that are often limited by smaller sample sizes. Moreover, our computational model uniquely isolates the dynamics of different internal organs and associates them with brain activities, both jointly and separately, an analysis that is not feasible in traditional experimental studies. This capability underscores the power of computational modeling, combined with a large-scale dataset, to advance knowledge in interoception. Meanwhile, our results indicate that several brain regions encode both respiratory and cardiac dynamics. This finding aligns with an emerging perspective based on high-precision fMRI that challenges the classic view of topographic homunculus representation of the bodily parts in the primary motor cortex [10].

Our model is limited to the ascending pathway, which describes how the brain senses and predicts the body, and lacks a descending pathway to account for how the brain descends motor commands to regulate physiological states. Our model may also be enhanced by using a deep neural network to encode brain activities and then fusing the brain representation with bodily representations. Although this strategy is likely to account for more complex brain-body association, it would preclude direct voxel-wise encoding and thus compromise direct functional mapping and explainability.

Nevertheless, the results from this initial work are encouraging, meriting future studies to scale up models with deeper architectures, more learnable parameters, and training with larger datasets, as well as testing for a diverse set of downstream tasks. Of particular interest would be using joint brain-body representations to predict individual behaviors relevant to cognitive abilities and emotional characteristics.

6 Conclusion

In summary, we built a model to represent cardiac and respiratory signals, first separately and then jointly, as a probabilistic latent distribution, and trained the model to predict both physiological signals forward in time. We further associated these latent representations with brain fMRI signals, enabling body-to-brain prediction and identifying the subcortical and cortical substrates that support this prediction. This model is an initial attempt to substantiate predictive coding for interoception, providing a valuable tool to quantify brain-body alignment and map central representations of visceral organs, holistically.

Acknowledgments and Disclosure of Funding

The work is partly supported by funding from the NIH (AT011665, AG082204) and NSF (IIS 2112773).

References

[1] Lisa Feldman Barrett and W. Kyle Simmons. "Interoceptive predictions in the brain". In: *Nature Reviews Neuroscience* 16.7 (2015). Publisher: Nature Publishing Group, pp. 419–429. ISSN: 1471-0048. DOI: 10.1038/nrn3950.

[2] Tom B. Brown et al. *Language Models are Few-Shot Learners*. 2020. DOI: 10.48550/arXiv.2005.14165. arXiv: 2005.14165[cs].

[3] Josue Ortega Caro et al. "BrainLM: A foundation model for brain activity recordings". In: *The Twelfth International Conference on Learning Representations*. 2023.

[4] Kyunghyun Cho et al. *Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation*. 2014. DOI: 10.48550/arXiv.1406.1078. arXiv: 1406.1078[cs].

[5] Jacob Devlin et al. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. 2019. DOI: 10.48550/arXiv.1810.04805. arXiv: 1810.04805[cs].

[6] Zijian Dong et al. *Brain-JEPA: Brain Dynamics Foundation Model with Gradient Positioning and Spatiotemporal Masking*. arXiv.org. 2024.

[7] Alexey Dosovitskiy et al. *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*. 2021. DOI: 10.48550/arXiv.2010.11929. arXiv: 2010.11929[cs].

[8] Tahnée Engelen, Marco Solcà, and Catherine Tallon-Baudry. "Interoceptive rhythms in the brain". In: *Nature Neuroscience* 26.10 (2023), pp. 1670–1684. ISSN: 1546-1726. DOI: 10.1038/s41593-023-01425-1.

[9] Matthew F. Glasser et al. "The minimal preprocessing pipelines for the Human Connectome Project". In: *NeuroImage* 80 (2013), pp. 105–124. ISSN: 1095-9572. DOI: 10.1016/j.neuroimage.2013.04.127.

[10] Evan M Gordon et al. "A somato-cognitive action network alternates with effector regions in motor cortex". In: *Nature* 617.7960 (2023), pp. 351–359.

[11] Irina Higgins et al. "beta-vae: Learning basic visual concepts with a constrained variational framework". In: *International conference on learning representations*. 2017.

[12] Musashi Hinck et al. *LLaVA-Gemma: Accelerating Multimodal Foundation Models with a Compact Language Model*. 2024. DOI: 10.48550/arXiv.2404.01331. arXiv: 2404.01331[cs].

[13] Jung-Hoon Kim et al. "Representation learning of resting state fMRI with variational autoencoder". In: *NeuroImage* 241 (2021), p. 118423. ISSN: 1095-9572. DOI: 10.1016/j.neuroimage.2021.118423.

[14] Diederik P. Kingma and Max Welling. *Auto-Encoding Variational Bayes*. 2022. DOI: 10.48550/arXiv.1312.6114. arXiv: 1312.6114[stat].

[15] Zhongming Liu et al. "4 - Interoceptive influences on resting-state fMRI". In: *Advances in Resting-State Functional MRI*. Ed. by Jean Chen and Catie Chang. Neuroimaging Methods and Applications. Academic Press, 2023, pp. 87–105. ISBN: 978-0-323-91688-2. DOI: 10.1016/B978-0-323-91688-2.00015-1.

[16] Jun Ma et al. “Segment anything in medical images”. In: *Nature Communications* 15.1 (2024). Publisher: Nature Publishing Group, p. 654. ISSN: 2041-1723. DOI: 10.1038/s41467-024-44824-z.

[17] George Mathew et al. “Foundation models for cardiovascular disease detection via biosignals from digital stethoscopes”. In: *npj Cardiovascular Health* 1.1 (2024). Publisher: Nature Publishing Group, p. 25. ISSN: 2948-2836. DOI: 10.1038/s44325-024-00027-5.

[18] Girish Narayanswamy et al. “Scaling Wearable Foundation Models”. In: *The Thirteenth International Conference on Learning Representations*. 2024.

[19] Alec Radford et al. *Learning Transferable Visual Models From Natural Language Supervision*. 2021. DOI: 10.48550/arXiv.2103.00020. arXiv: 2103.00020[cs].

[20] Nikhila Ravi et al. *SAM 2: Segment Anything in Images and Videos*. 2024. DOI: 10.48550/arXiv.2408.00714. arXiv: 2408.00714[cs].

[21] Anil K. Seth. “Interoceptive inference, emotion, and the embodied self”. In: *Trends in Cognitive Sciences* 17.11 (2013), pp. 565–573. ISSN: 1879-307X. DOI: 10.1016/j.tics.2013.09.007.

[22] Anil K. Seth, Keisuke Suzuki, and Hugo D. Critchley. “An interoceptive predictive coding model of conscious presence”. In: *Frontiers in Psychology* 2 (2011), p. 395. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2011.00395.

[23] David C. Van Essen et al. “The WU-Minn Human Connectome Project: An overview”. In: *NeuroImage*. Mapping the Connectome 80 (2013), pp. 62–79. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2013.05.041.

[24] Fan Yang et al. “scBERT as a large-scale pretrained deep language model for cell type annotation of single-cell RNA-seq data”. In: *Nature Machine Intelligence* 4.10 (2022). Publisher: Nature Publishing Group, pp. 852–866. ISSN: 2522-5839. DOI: 10.1038/s42256-022-00534-z.

A Technical Appendices and Supplementary Material

A.1 Interoception

The brain senses, integrates, predicts, and regulates bodily states through bidirectional communication with the internal organs. This process is known as interoception. Beyond regulating physiological responses and maintaining homeostasis, interoception also influences our motivation and emotions and affects decisions and behaviors. Its functionality relies on both neuronal pathways in the peripheral and central nervous system, as well as the circulating hormones. In the central nervous system, several experimental studies have shown that interoceptive afferent signals are first processed by subcortical nuclei, such as the nucleus of the solitary tract (NTS), the parabrachial nucleus (PBN), the locus coeruleus (LC), and the ventral posteromedial nucleus of the thalamus (VPM). These signals are further integrated by the cortical areas. In particular, the insular cortex (IC) serves as a key sensory hub for integrating bodily signals and relaying them to higher-order brain regions, including the amygdala (AMY), orbitofrontal cortex (OFC), anterior cingulate cortex (ACC), and prefrontal cortex (PFC). Finally, these signals project back to the lower brainstem and convey efferent motor commands to visceral organs, allowing for top-down regulation of the internal organs.

Here, we provide a comprehensive list of brain regions mentioned in the paper and figure. ACC: anterior cingulate cortex, AIC: anterior insular cortex, dlPFC: dorsal lateral prefrontal cortex, DVC: dorsal vagal complex, IPC: inferior parietal cortex, MCC: mid cingulate cortex, M1: primary motor cortex, NG: nodose ganglion, PAC: primary auditory cortex, PIC: posterior insular cortex, PCC: posterior cingulate cortex, S1: primary sensory cortex, TPO: temporal-parietal-occipital junction, V1: primary visual cortex, MT: mid temporal area.

A.2 Canonical Correlation Analysis for Brain-Body Association

We perform CCA to map the interoceptive network. Specifically, we first linearly reconstruct the rs-fMRI time series and then estimate the Pearson correlation coefficients between the true and reconstructed time series. Below, we provide a mathematical derivation of the reconstruction of the rs-fMRI time series.

Take the vertices at the cortical surface as an example. We estimate their covariance with the context vector **c** as **COV** and perform SVD to get singular vectors **U** and **V**, and singular values Σ, following

$$\mathbf{COV} = \mathbf{U}\Sigma\mathbf{V}^T.$$

Assuming that **X** and **Y** are brain activity with the shape of (n_{voxels}, n_{times}) and bodily representations with the shape of (n_{dims}, n_{times}), respectively, we project **X** and **Y** to the shared space following

$$\mathbf{X}_{proj} = U[:, 0:n]^T\mathbf{X},$$

$$\mathbf{Y}_{proj} = \Sigma[0:n, 0:n]V[:, 0:n]^T\mathbf{Y},$$

where n is the number of singular vectors ($n = 4$) chosen empirically, and $\mathbf{X}_{proj}$ and $\mathbf{Y}_{proj}$ are projected signals. Given that two projected signals are in the same shared space, we can combine the above two equations into

$$U[:, 0:n]^T\mathbf{X} \approx \Sigma[0:n, 0:n]V[:, 0:n]^T\mathbf{Y}.$$

Its reorganization leads to

$$\mathbf{X} \approx U[:, 0:n]\Sigma[0:n, 0:n]V[:, 0:n]^T\mathbf{Y}.$$

Therefore, we can approximately reconstruct or predict the brain activity ($\hat{\mathbf{X}}$) based on its co-fluctuation with the bodily states, with the following equation

$$\hat{\mathbf{X}} = U[:, 0:n]\Sigma[0:n, 0:n]V[:, 0:n]^T\mathbf{Y}.$$

We calculate the Pearson correlation coefficients between the reconstructed brain activities and the true brain activities for all vertices on the cortical surface across all subjects in the test set. Brain regions with higher correlation coefficients fluctuate similarly with the physiological signals, thus better encoding bodily states.

A.3 Statistical Analysis

The singular vectors in $\mathbf{U}$ reveal the brain patterns that co-fluctuate with the physiological signals. We estimate and visualize the top four singular vectors from the cortical and subcortical spaces, separately. Specifically, we first calculate the covariance (1) between the context vectors $\mathbf{c}$ and every vertice in the cortical surface, or (2) between the context vectors $\mathbf{c}$ and every voxel in the subcortical space, followed by SVD to get singular vectors $\mathbf{U}$. Each element in a singular vector represents the value of a voxel, mapping its association with the bodily states. We perform a blocked permutation test to highlight voxels with associations that are higher than the chance level. To obtain the null distribution, we randomly permute the rs-fMRI test series 1,000 times, using a block size of 45 seconds, and repeat the procedure to estimate the singular vectors. The significance level (α) is 0.05. Additionally, we perform the Benjamini-Hochberg procedure to correct for multiple comparisons and control the false positive rate.

A.4 Organ-Specific Association with Brain

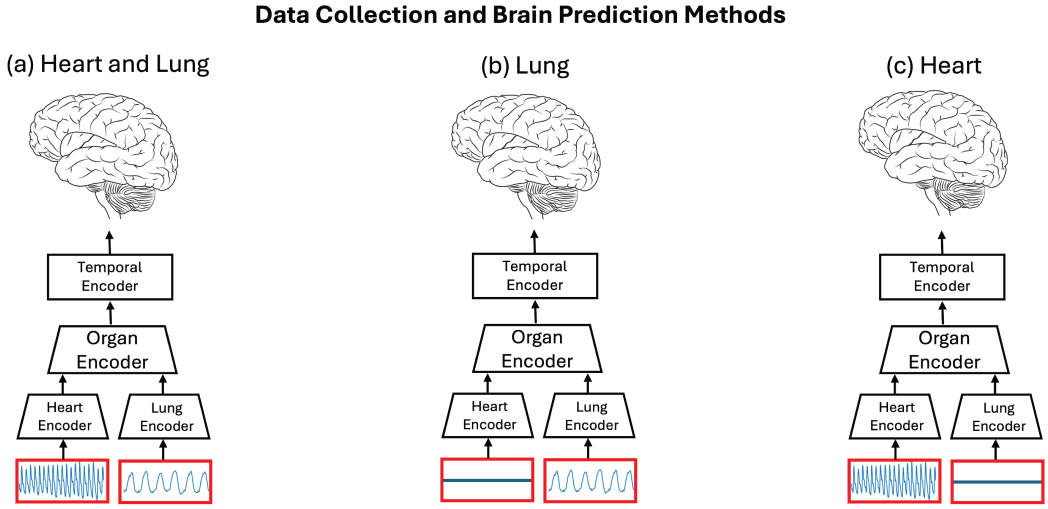

Figure 4: The framework to reconstruct or predict brain activities from context vectors. (a) Both cardiac input and respiratory input are provided to the model to extract the context vectors for reconstructing brain activities. (b) Only respiratory input is provided to the model to extract the context vectors for reconstructing brain activities. Zero values are used to replace the cardiac input. (c) Only cardiac input is provided to the model to extract the context vectors for reconstructing brain activities. Zero values are used to replace the respiratory input. Using this framework, we estimate the interoceptive network that encodes both cardiac and respiratory input from (a), only respiratory input from (b), and only cardiac input from (c).

We isolate the organ-specific contribution to the interoceptive network and explore whether organ-specific mapping exists in the brain. Fig. 4 details the method. For example, we isolate the brain regions with activities that only fluctuate with the respiratory signals by silencing the cardiac input (Fig. 4b). The model takes segments of the true respiratory trace and fake cardiac input (zero values) and outputs corresponding context vectors $\mathbf{c}$. We repeat the CCA (detailed in A.2) between these context vectors $\mathbf{c}$ and rs-fMRI signals in the cortical space. In this way, we map the brain regions with activities that fluctuate only with the respiratory signal, but not with the cardiac signals. A

similar approach can isolate the brain regions with activities that fluctuate only in response to cardiac signals. Taken together, we leverage the power of the proposed model to map the organ-specific and organ-sharing brain regions in the interoceptive network.

A.5 Supplementary Results

A.5.1 Common Space for Brain and Body Fluctuations

Brain activities fluctuate with bodily states in shared space. The top several singular vectors, estimated from the covariance matrix between the two, span the shared space. In this study, we select four singular vectors to form the axis of the shared space. In particular, the first singular vectors $\mathbf{u}_1$ and $\mathbf{v}_1$, along which the most aligned activities between the brain and the body are observed. The alignments between projected signals along the first singular vectors are shown in Fig. 5. Note that the analysis at the cortical and subcortical levels is performed separately. We independently calculate the covariance matrices for cortical signals and subcortical signals, and estimate their singular vectors to yield two separate shared spaces related to bodily states. However, the projected bodily states in two spaces align well with each other, and so do the projected cortical and subcortical activities. This observation shows that these two shared spaces highly agree with each other, potentially indicating a three-way co-fluctuation between the cortical, subcortical activities, and bodily states.

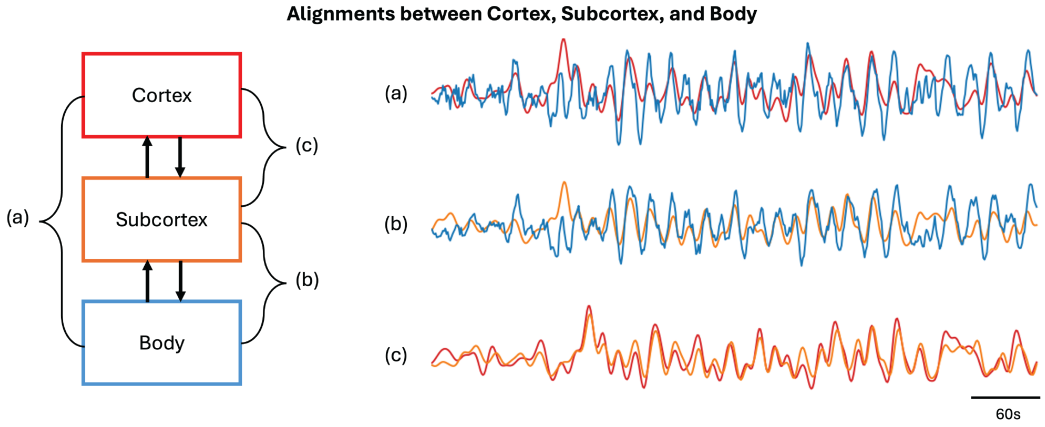

Figure 5: Projections of rs-fMRI and physiological signals to the shared space spanned by the first singular vectors. Signals from the brain and the body are better aligned in the shared space. (a) Alignment between the projected cortical activities and bodily states (context vectors), (b) Alignment between the projected subcortical activities and bodily states (context vectors), (c) Alignment between the projected cortical and subcortical activities.

A.5.2 Cortical Brain Maps

We visualize the top four singular vectors $\mathbf{u}_1$ to $\mathbf{u}_4$ in the cortical space in Fig.6. The patterns highlight the insular cortex, the somatosensory cortex, and the visual and auditory cortex. Compared with $\mathbf{u}_1$ and $\mathbf{u}_2$, $\mathbf{u}_3$ and $\mathbf{u}_4$ highlight more find-grained brain regions.

A.5.3 Subcortical Brain Maps

We visualize the first singular vectors $\mathbf{u}_1$ in the subcortical space in Fig.7. The patterns highlight the thalamus, hippocampus, and brainstem.

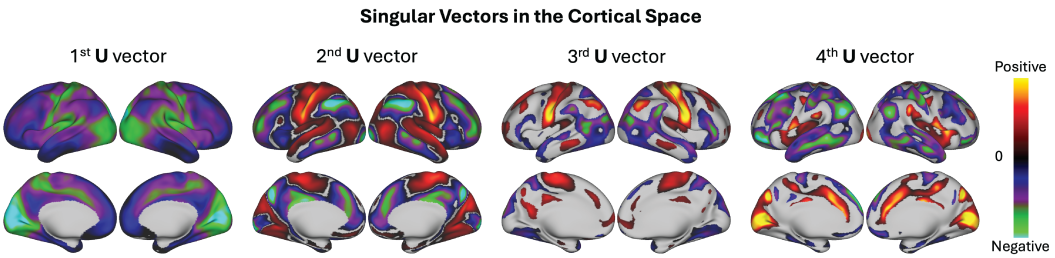

Figure 6: The top four singular vectors in $\mathbf{U}$ are visualized in the cortical space. Only vertices with a statistically significant difference from the null distribution are displayed ($p < 0.05$).

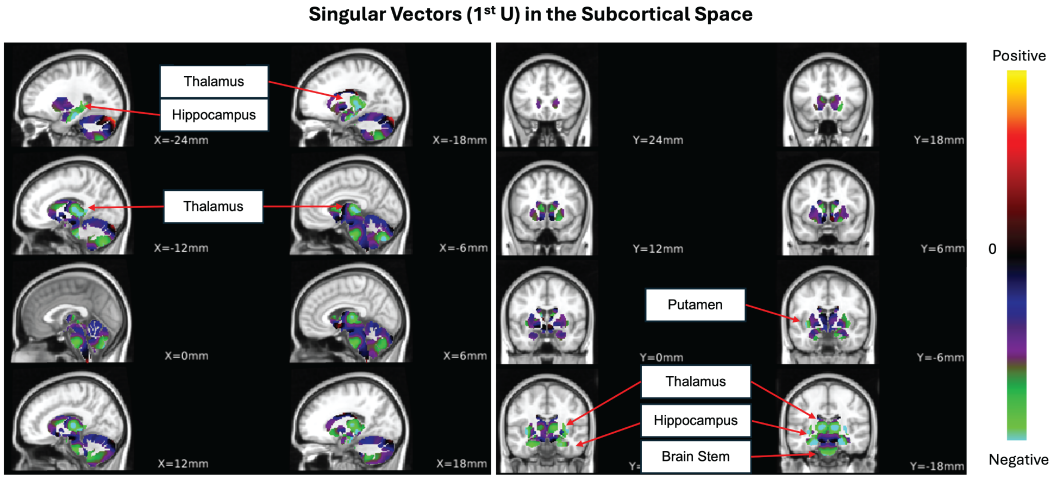

Figure 7: The first singular vector in $\mathbf{U}$ is visualized in the subcortical space. Only voxels with a statistically significant difference from the null distribution are displayed ($p < 0.05$).