

Neurocognitive Latent Space Regularization for Multi-Label Diagnosis from MRI

Jocasta Manasseh-Lewis¹[0009–0003–3823–6908]*, Felipe Godoy¹[0009–0007–7332–5016]*, Wei Peng¹[0000–0002–2892–5764], Robert Paul², Ehsan Adeli¹[0000–0002–0579–7763]**, and Kilian Pohl¹[0000–0001–5416–5159]

¹ Stanford University, Stanford, CA 94305

² University of Missouri – St. Louis, St. Louis, MO 63121

Abstract. Interpretability is essential to MRI brain studies relying on deep learning models for neuroscientific discovery. One way to facilitate the interpretability of a deep learning model is to ensure the samples are arranged in the model’s latent space with respect to clinically meaningful variables. To achieve this in the context of cross-sectional brain MRI studies, we regularize the latent space of a multi-label classifier via pairwise disentanglement, so that the difference between the representation of two brain MRIs along the disentangled direction in the latent space is similar to the difference in their neuropsychological test scores. We apply our technique to classify brain MRIs of 156 controls, 165 cases diagnosed with mild cognitive impairment (MCI), 166 diagnosed with human immunodeficiency virus (HIV)-associated cognitive disorder (HAND), and 32 individuals diagnosed with HIV without HAND. The latent space is disentangled with respect to the neuropsychological z-score (NPZ), which is negatively correlated with the severity of cognitive impairment (i.e., low scores for those diagnosed with MCI or HAND). Based on cross-validation, the proposed model achieves statistically significantly higher balanced accuracy than the same model without disentanglement. Furthermore, the difference between representations along the disentangled direction significantly correlates with the difference in NPZ. Finally, the brain regions guiding the classification process aligned with the neuroscientific literature.

Keywords: disentanglement, pairwise learning, multi-label MRI classification, cognitive impairment, HIV

1 Introduction

Recent deep learning research has shown that steering representation learning paradigms toward the discovery of interpretable features can enhance model prediction accuracy, while improving generalization, robustness, and providing

* Authors contributed equally

** Corresponding author: eadeli@stanford.edu

valuable insights into the task [4,34]. An example is disentanglement-based representation learning, which enforces the learning of a latent space holding interpretable properties relative to hidden factors present in a dataset [35]. With respect to brain MRI datasets, disentanglement has been commonly used to extract time-varying factors such as brain age in longitudinal contexts [15,22,32,37], modality-dependent information in multi-modal settings [27,11], and the underlying MRI morphological features (shape and appearance) [5].

These previous approaches, involving time-varying or morphological hidden factors, typically isolate underlying features for each subject [35]. However, to advance neuroscientific understanding and develop interpretable latent representations in cross-sectional studies, it is essential to establish a manifold over which the subjects are ordered relative to each other, so that the individual samples can be positioned consistently along a shared spectrum [13]. To achieve this, we propose a method based on pairwise sampling during training. By utilizing a clinically meaningful continuous variable (such as cognitive test scores) as the hidden factor, we can create a disentangled representation space. This space is then used for multi-class cross-sectional classification, providing a more unified and interpretable framework for analyzing the data. The training of the classifier is done jointly with disentangling the hidden factor so that the disentanglement task can be interpreted as a model regularizer [9].

In this paper, we test our model on an MRI dataset consisting of 519 cross-sectional brain MRIs from 156 healthy controls, 165 individuals suffering from mild cognitive impairment (MCI), 32 people living with human immunodeficiency virus without cognitive impairment (HIV), and 166 individuals diagnosed with HIV-Associated Neurocognitive Disorder (HAND). The data also includes the NPZ-global score [20], a common score for assessing neuropsychological functioning in HIV-positive individuals, which we use in our experiments as the hidden factor. On this dataset, cross-validation reveals that performing the disentanglement when identifying these four cohorts results in interpretable features that improve classification accuracy.

2 Methodology

To improve multi-label classification of 3D MRIs, we propose a pairwise representation learning framework disentangling the latent-space with respect to a continuous factor related to the classification task. Specifically, let each sample in a dataset be represented by a structural 3D-MRI x and a continuous variable σ , such as a cognitive test score. The 3D-MRI x is the input to an encoder, whose latent space will be disentangled with respect to σ (the hidden factor). The latent representation produced by the encoder is then the input to the linear classifier predicting diagnosis. We train the model so that it jointly optimizes the binary cross-entropy (BCE) loss related to the classification task and a pairwise representation loss specific to the disentanglement. In other words, the representation loss acts as a regularizer to the classification task during training.

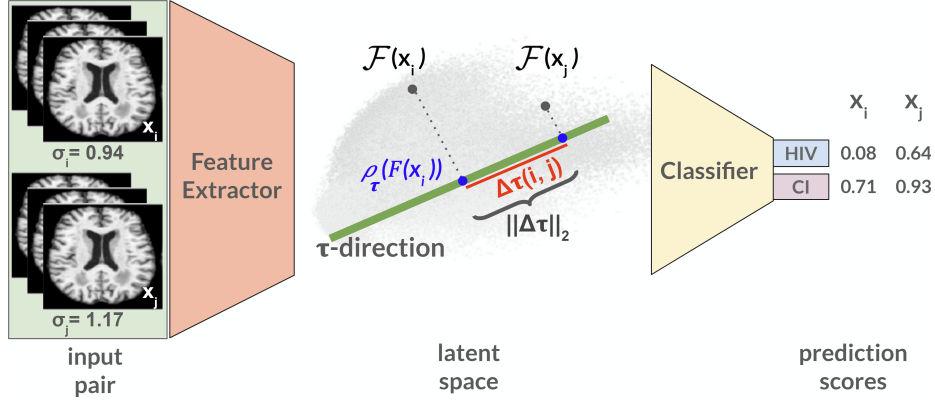


Fig. 1: Our proposed pairwise disentanglement setup. The image embeddings on the latent space $\mathcal{F}(x_i), \mathcal{F}(x_j)$ are projected onto the linear latent-direction τ , giving the points $\rho_\tau(\mathcal{F}(x_i)), \rho_\tau(\mathcal{F}(x_j))$. The disentanglement loss function encourages the distance between the projected points $\Delta\tau(i, j)$ to be similar to that of the hidden factor $\Delta\sigma(i, j)$.

The goal of the disentanglement is to learn a linear direction τ in the latent space that is associated with the values of the hidden factor σ in the input space (see Figure 1). Specifically, let $\mathcal{D}$ be the space of all possible pairing of samples of the training set and $(i, j) \in \mathcal{D}$ be a random sample from that space. (σ_i, σ_j) is the corresponding pair of hidden factor values, (x_i, x_j) the pairs of MRIs, and $(\mathcal{F}(x_i), \mathcal{F}(x_j))$ the corresponding embeddings in the latent space produced by the encoder. Let $\Delta\sigma(i, j) := \sigma_i - \sigma_j$ be the difference between their hidden values and $\Delta\tau(i, j)$ be the difference between $(\mathcal{F}(x_i), \mathcal{F}(x_j))$ projected onto τ . Let $\rho_\tau(\cdot)$ be the projection operator onto τ and τ be a linear direction, then $\Delta\tau(i, j) = \rho_\tau(\mathcal{F}(x_i)) - \rho_\tau(\mathcal{F}(x_j))$. To encourage a linear relationship between $\Delta\sigma$ and $\Delta\tau$, we define the loss function for the disentanglement task to be

$$\begin{aligned}
L_r &:= \sum_{(i,j) \in \mathcal{D}} (||\Delta\tau(i, j)||_2 - |\Delta\sigma(i, j)|) \\
&= \sum_{(i,j) \in \mathcal{D}} (||\rho_\tau(\mathcal{F}(x_i)) - \rho_\tau(\mathcal{F}(x_j))||_2 - |\sigma_i - \sigma_j|),
\end{aligned}$$

where $||\cdot||_2$ is the L-2 norm operator. Note, this cost function could easily be extended to τ being an arbitrary manifold. However, constraining τ to be linear provides an intuitive interpretation of a direction in the latent space and reduce the computational burden.

In our model τ is a learnable parameter defined separately from the encoder and classifier, so the optimal τ direction to associate with σ can be determined in a self-supervised manner during training. In that way the loss function L_r relates to the self-supervised disentanglement approaches [22, 39]. However, these self-supervised models are confined to tracking changes over time by confining pairs

Measure	CTRL	HIV	MCI	HAND	Comparison
Number of samples	156	32	165	166	-
Sex (F/M)	10/146	0/32	84/81	11/155	$p < 0.01$
Age (years)	70.2 ± 5.87	64.7 ± 3.42	67.7 ± 7.80	64.3 ± 4.92	HAND=HIV<MCI<CTRL
Ethnicity (A/AA/C/O)	13/1/139/2*	0/0/32/0	12/0/38/0*	4/16/117/12*	$p = 0.7$
NPZ (n)	0.22 ± 0.55 (113)	0.10 ± 0.42 (21)	-0.80 ± 0.94 (85)	-0.26 ± 0.25 (29)	MCI<HAND<HIV=CTRL
Education (yrs)	16.3 ± 3.25	16.9 ± 2.31	16.0 ± 2.24	17.6 ± 2.08	MCI=CTRL=HIV<HAND
Handedness (R/L/A)	103/12/2*	26/6/0	120/20/4*	136/13/0*	$p = 0.8$

Table 1: Demographic of UCSF dataset. ‘*’ indicates missing values and n indicates the number of cohort samples with known NPZ score. A/AA/C/O in Ethnicity represents Asian/African American/Caucasian/Other, and R/L/A in Handedness represents right-handed/left-handed/ambidextrous. Multiple one sided t-tests and ANOVA were performed to determine significance differences in per-class demographics.

to be composed of two assessments from the same subject. Thus, these methods are only suitable for longitudinal studies. Unlike them, our cross-sectional model can track changes in the latent space with respect to any continuous value. Finally, let L_c be the cost function of the classifier defined by binary cross entropy, then the total loss function is defined as $\mathcal{L} = L_c + \lambda L_r$.

3 Experimental Setup

Dataset We apply our method to a dataset collected by the Memory and Aging Center of the University of California, San Francisco (UCSF, PI: Anonymous). As detailed in Table 1, the UCSF dataset consisted of T1-weighted MRIs ($138 \times 176 \times 138$ voxels) from 156 controls, 165 MCIs, 32 HIVs with regular cognition, and 166 HAND patients. Most subjects also underwent a sixty-minute neuropsychological assessment covering domains needed to identify cognitive impairment, including a finger tapping (FT) test to assess their motor control integrity of the neuromuscular system, following published criteria [3]. The test was summarized by the Neuropsychological Z-global Score (NPZ score) [20]. Diagnosis of HAND and MCI were determined at consensus conference following [1,2]. HIV-positive status was defined based on reacting HIV antibody test.

Minimal processing of each MRI consists of bias field correction using the Computational Morphometry Toolkit (CMTK) [29], skull stripping using majority voting based on Robust Brain Extraction (ROBEX) [18], FSL BET [21], and AFNI [12]. The resulting skull stripped MRI is affinely registered to the SRI24 atlas [30] and resized to $64 \times 64 \times 64$ voxels.

Model Architecture The fixed model architecture builds upon [38]. Specifically, we use a convolutional neural network (CNN) to perform feature extraction on 3D MRI brain scans through multiple 3D convolutional (Conv-3D) layers. The encoder component consists of four $3 \times 3 \times 3$ Conv-3D layers with 16, 32, 64, and 32 output channels respectively, rectified linear (ReLU) activation, 3D

Model	CTRL	MCI	HIV	HAND	Overall
Baseline	33.63% $\pm$ 6.61%	57.61% $\pm$ 5.61%	17.48% $\pm$ 5.45%	50.88% $\pm$ 6.92%	39.90% $\pm$ 1.27%
Reg-Dis	38.53% $\pm$ 9.83%	57.85% $\pm$ 6.38%	31.03% $\pm$ 10.15%	54.70% $\pm$ 6.97%	45.51% $\pm$ 3.32%

Table 2: Mean and standard deviation in balanced accuracy for baseline and Reg-Dis ($\lambda = 0.25$) with respect to each cohort and overall. The overall balanced accuracy is significantly higher ($p=0.009$) for Reg-Dis.

batch normalization [19], and 3D max-pooling layers with $2 \times 2 \times 2$ kernels. The output of the encoder is a 2048D-vector, whose corresponding latent space was disentangled with respect to the NPZ score.

The resulting vector representation is then passed through a classifier to yield a two-dimensional prediction output: one for HIV status and one for cognitive impairment (CI) status. Thus, HAND patients are correctly labeled with positive HIV and CI status. The classifier is composed of four fully connected layers, reducing the vector dimensionality in the sequence $2048 \rightarrow 256 \rightarrow 128 \rightarrow 16 \rightarrow 2$. The linear connections include dropout operations with probability parameter 0.25 and leaky rectified linear activation (leaky ReLU). At the prediction step, the 2D output is applied to a Sigmoid activation function to yield an output $\in [0, 1]^2$. The implementation is trained using a single instance of an NVIDIA GeForce RTX 2080Ti GPU with 12 GB of RAM.

Model Evaluation Using five-fold stratified cross-validation, we assess the balanced accuracy of the model over the folds. During each training, the random pairing of subjects and calculation of the disentanglement loss is confined to those with recorded NPZ scores. As the diseases under investigation are not known to impact one hemisphere more than the other, we randomly flip MRIs across the sagittal plane (so that the left hemisphere ends up on the right and vice versa) and apply affine transformation to generate a total of 1080 augmented training samples. We then pick the optimal hyperparameter λ from the search space $\{0.01, 0.05, 0.1, 0.25, 0.5, 1.0\}$ whose balanced accuracy on the training set is highest. The corresponding model is then applied to the test set to record its balanced accuracy. We refer to this outcome as regularized-disentanglement (Reg-Dis). To better understand the impact of the regularized disentanglement, we create a baseline by repeating the analysis fixing $\lambda = 0$, which instantiates a model without disentanglement.

Saliency We compute the average saliency map for all MRIs in the test set by Reg-Dis via back-propagation. The classifier produces the two 'raw' saliency maps s_{CI} and s_{HIV} for all test subjects. To highlight areas of the MRI most important for classification, the function $\max(s_{CI} - s_{HIV}, 0)$ for MCI and $\max(s_{HIV} - s_{CI}, 0)$ for HIV is applied. For HAND, the two saliency maps are averaged together, i.e., $(s_{MCI} + s_{HIV}) / 2$. Each of the three saliency maps are averaged

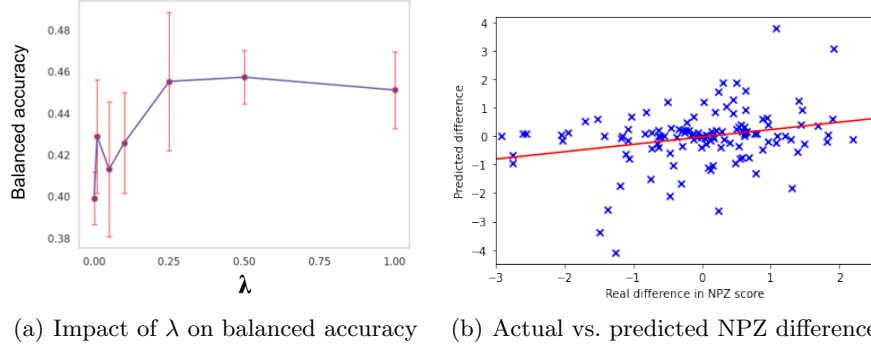


Fig. 2: (a) the balanced accuracy of Reg-Dis with respect to λ reveals that the approach is relatively stable for $\lambda \geq 0.25$. (b) the difference in NPZ score between random pairs of samples reveals a statistically significant correlation ($p < 0.01$) between the actual and predicted difference. The predicted difference is the distance of the two samples projected on τ in the latent space.

across the left and right hemispheres. Furthermore, they are normalized by the 95th quantile of their saliency values and thresholded to exclude voxels with z-score < 2 . Finally, each outcome is superimposed onto an atlas in order to identify the brain regions most important in the decision process.

4 Quantitative Evaluation

The overall balanced accuracy of Reg-Dis is 46.13 %, which is significantly higher ($p = 0.009$; Wilcoxon Rank Sum test) than the baseline (39.9%). However, the choice of λ impacts the accuracy of the model according to Fig. 2(a) with $\lambda < 0.25$ not emphasizing the disentanglement enough but then becoming stable (for $\lambda \geq 0.25$) and with slight over-regularization for $\lambda > 0.5$. Interestingly, the least difference between the per-class accuracy scores (computed as recall rate) across the cohorts is achieved with $\lambda = 0.25$ (Table 2). The balanced accuracy of the Reg-Dis model (45.5%) is 5.6% higher than the baseline model (39.9%), which is significant according to the Wilcoxon Rank Sum test ($p = 0.009$). The greatest improvement for per-class accuracy is observed for the HIV-only cohort, which is lower than random-guessing (i.e., 25%) for the baseline model and improves to 31% for Reg-Dis.

More importantly, only the predictions by the Reg-Dis model are not confounded with respect to demographic factors which are significantly different between the cohorts (Table 1), those being age, education, and sex. For each prediction model and status (i.e., CI or HIV), we test the predictions (before being applied the sigmoid function) generated on the test folds for all CTRLs via generalized linear model (GLM) (see also [28]). The predictors of each GLM are

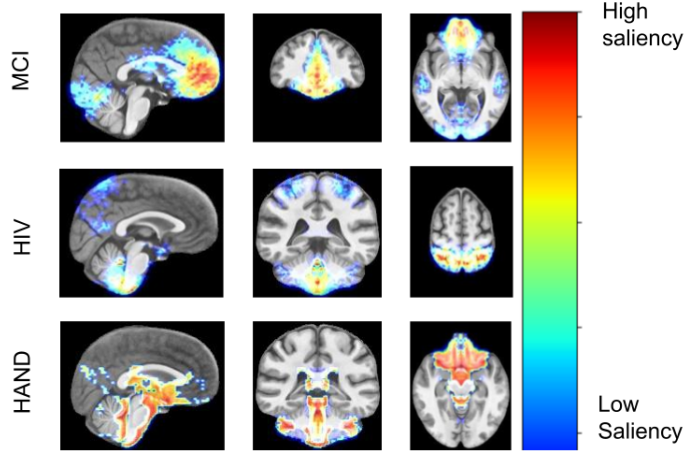


Fig. 3: Saliency maps for MCI, HIV, and HAND. Red are brain regions of high importance for the classification while blue indicates low importance. For MCI, regions of importance are the superior temporal gyrus, caudate nucleus, middle occipital gyrus, olfactory and lingual cortex. For HIV, it is the superior and middle frontal gyrus, pars triangularis, cerebellar white matter, precuneus, and posterior cingulate cortex. Finally for HAND, the medial and superior temporal pole, cerebellar white matter, superior and inferior orbitofrontal cortex, and parahippocampal gyrus are important for prediction.

the possible confounders (i.e., age, education and sex). This analysis identifies significant association between CI and sex for the baseline but none for Reg-Dis.

To demonstrate that the relationship between the NPZ score and the values on τ are linear, we randomly pair subjects within their respective test folds, embed them onto the latent-space according to the trained model, and apply the disentanglement operation described in Section 2. Next, we compute the Pearson correlation between the difference with respect to σ and the difference in τ . This correlation ($r = 0.27$) was significant ($p < 0.01$) which Figure 2(b) confirms.

5 Regions Driving Predictions

Figure 3 shows the saliency maps for MCI, HIV (with normal cognition), and HAND. The most important regions for identifying MCI were the superior temporal gyrus, caudate nucleus, middle occipital gyrus, olfactory and lingual cortex. Regions important for identifying HIV (with normal cognition) were the frontal gyrus (pars triangularis, middle, and superior), precuneus, posterior cingulate cortex, and the cerebellar white matter. The saliency in the cerebellar white matter was even higher for identifying HAND. Unique to HAND were

high salience in the superior and medial temporal pole, parahippocampal gyrus, and the superior and inferior orbitofrontal cortex.

Many of these brain regions are impacted by those diseases according to the literature. For example, olfactory dysfunction precedes other symptoms in the Alzheimer’s continuum, including MCI [24], as it is a early site of amyloid beta plaque deposition and neuroinflammation due to exposure to environmental toxic factors [33]. Furthermore, olfactory impairment is particularly severe in amnesic MCI and can be used to accurately screen for MCI [40]. Average thickness of the temporal gyrus is also found to be associated with MCI [17]. The superior temporal gyrus plays a role in speech perception [14], which is impacted by MCI [23]. Grey matter volume of the caudate nuclei was positively correlated with the degree of social intelligence [26], and social deficits [10] can be clinically observed in individuals with MCI.

HIV, on the other hand, is linked to changes in cortical thickness in the frontal lobes and cingulate gyrus [31,6]. Interestingly, subjective symptoms of neuropathy (pain, aching, or burning, sensation of pins and needles, or numbness in legs or feet) in HIV-positive individuals are associated with smaller precuneus volume, whereas objective symptoms (deep tendon ankle reflex, vibration sense, position sense, and two-point discrimination) are associated with cerebellar volume [36]. Cerebellar white matter is not only important for classifying HIV but also for HAND, which is not surprising given that cerebellum plays a critical role for motion and cognition [7]. With respect to HAND, widespread brain atrophy is reported, including for the orbitofrontal cortex, frontal and temporal lobes, and cerebellum [8]. Decreased volume in the parahippocampal gyrus, another region important for identifying HAND, is associated with neurological soft signs [16] in HIV positive individuals. The cerebellum, precuneus, parahippocampal gyrus, and superior temporal pole are all HIV-salient regions identified by Reg-Dis that were identified in [25] as differentiating regions between HAND and MCI. Notably, MCI is defined by six regions of possible deficit, represented by some of the regions Reg-Dis identifies as salient: learning and memory (olfactory cortex), language, complex attention, executive function (both preceding STG), social cognition (caudate, rSTG), and visuospatial function (occipital gyrus) [7] but does not typically include motor deficits as supported by the exclusion of the cerebellum in the saliency map.

6 Conclusion

We proposed a pairwise disentanglement method of regularization in the context of classifying 3D MRIs by diagnosis. We demonstrated how our technique can be applied to cross-sectional brain MRIs by regularizing the latent space of the model with respect to a cognitive score relevant to the analysis. Doing so improves the balanced accuracy of the classifier. Furthermore, disentanglement resulted in a more interpretable latent space as the disentangled direction had a linear relationship with the hidden factor. Moreover, the saliency maps highlighted brain regions impacted by MCI, HIV (without cognitive impairment),

and HAND that are aligned with the neuroscientific literature, demonstrating that creating more interpretable deep learning methods can translate also into neuroscientific insights.

Acknowledgement The work was partly funded by the U.S. National Institutes of Health (NIH) grants (AA017347, MH113406, AG073104, AG072701), Stanford HAI Google Cloud Credit, the DGIST Joint Research Project, and the 2024 Stanford HAI Hoffman-Yee Grant.

References

1. Albert, M.S., DeKosky, S.T., Dickson, D., Dubois, B., Feldman, H.H., Fox, N.C., Gamst, A., Holtzman, D.M., Jagust, W.J., Petersen, R.C., Snyder, P.J., Carrillo, M.C., Thies, B., Phelps, C.H.: The diagnosis of mild cognitive impairment due to Alzheimer’s disease: Recommendations from the national institute on aging-Alzheimer’s association workgroups on diagnostic guidelines for Alzheimer’s disease. *Focus* **11**(1), 96–106 (2013)
2. Antinori, A., Arendt, G., Becker, J.T., Brew, B.J., Byrd, D.A., Cherner, M., Clifford, D.B., Cinque, P., Epstein, L.G., Goodkin, K., Gisslen, M., Grant, I., Heaton, R.K., Joseph, J., Marder, K., Marra, C.M., McArthur, J.C., Nunn, M., Price, R.W., Pulliam, L., Robertson, K.R., Sacktor, N., Valcour, V., Wojna, V.E.: Updated research nosology for HIV-associated neurocognitive disorders. *Neurology* **69**(18), 1789–1799 (2007)
3. Axelrod, B.N., Meyers, J.E., Davis, J.J.: Finger tapping test performance as a measure of performance validity. *The Clinical Neuropsychologist* **28**(5), 876–888 (2014)
4. Bengio, Y., Courville, A., Vincent, P.: Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35**(8), 1798–1828 (2013)
5. Bercea, C.I., Wiestler, B., Rueckert, D., Albarqouni, S.: Federated disentangled representation learning for unsupervised brain anomaly detection. *Nature Machine Intelligence* **4**(8), 685–695 (2022)
6. Cabral, R.F., Diogo Goulart Corrêa and, N.Z., Tukamoto, G., Kubo, T.T.A., Fonseca, R.P., Silva, M.M., Wilner, N.V., Bahia, P.R.V., Gasparetto, E.L., Marchiori, E.: Preliminary comparative study of cortical thickness in HIV-infected patients with and without working memory deficit. *PloS One* **16**, e0261208 (2021)
7. Cao, S., Nie, J., Zhang, J., Chen, C., Wang, X., Liu, Y., Mo, Y., Du, B., Hu, Y., Tian, Y., Wei, Q., Wang, K.: Role of fronto-limbic circuit in neuropsychiatric symptoms of dementia: clinical evidence from an exploratory study. *Frontiers in Psychiatry* **15** (2024)
8. Chang, L., Shukla, D.K.: Chapter 18 - imaging studies of the HIV-infected brain, handbook of clinical neurology. *The Neurology of HIV Infection* **152**, 229–264 (2018)
9. Chen, X., Kingma, D.P., Salimans, T., Duan, Y., Dhariwal, P., Schulman, J., Sutskever, I., Abbeel, P.: Variational lossy autoencoder. *arXiv preprint arXiv:1611.02731* (2016)

10. Chen, Y., He, H., Ding, Y., Tao, W., Guan, Q., Krueger, F.: Connectome-based prediction of decreased trust propensity in older adults with mild cognitive impairment: A resting-state functional magnetic resonance imaging study. *NeuroImage* **292**, 120605 (2024)
11. Cho, J., Liu, X., Xing, F., Ouyang, J., Fakhri, G.E., Park, J., Woo, J.: Disentangled multimodal brain MR image translation via transformer-based modality infuser. In: *Medical Imaging 2024: Image Processing*. vol. 12926, p. 129262H (2024)
12. Cox, R.W.: Afni: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research* **29**(3), 162–173 (1996)
13. Fumero, M., Wenzel, F., Zancato, L., Achille, A., Rodolà, E., Soatto, S., Schölkopf, B., Locatello, F.: Leveraging sparse and shared feature activations for disentangled representation learning **36**, 27682–27698 (2023)
14. Han, J.H., Kim, J.H., Park, G.K., Lee, H.J.: Preserved gray matter volume in the left superior temporal gyrus underpins speech-in-noise processing in middle-aged adults. *The Journal of International Advanced Otology* **20**, 62–68 (2024)
15. Harada, S., Bise, R., Hayashi, H., Tanaka, K., Uchida, S.: Order-guided disentangled representation learning for ulcerative colitis classification with limited labels. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*. MICCAI 2021. *Lecture Notes in Computer Science* **12902**, 471–480 (2021)
16. Herold, C.J., Kong, L., Ceballos, M.E., Schröder, J., Toro, P.: Neurological soft signs and brain morphology in people living with HIV. *Journal of NeuroVirology* **28**, 236–247 (2022)
17. Hojjati, S.H., Abbas Babajani-Feremi, f.t.A.D.N.I.: Seeing beyond the symptoms: biomarkers and brain regions linked to cognitive decline in Alzheimer’s disease. *Frontiers in Aging Neuroscience* **16**, 1356656 (2024)
18. Iglesias, J.E., Liu, C.Y., Thompson, P.M., Tu, Z.: Robust brain extraction across datasets and comparison with publicly available methods. *IEEE Transactions on Medical Imaging* **30**, 1617–1634 (2011)
19. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: *Proceedings of the 32nd International Conference on Machine Learning*. pp. 448–456 (2015)
20. Janssen, R.S., Cornblath, D.R., Epstein, L.G., Foa, R.P., et al.: Nomenclature and research case definitions for neurologic manifestations of human immunodeficiency virus-type 1 (HIV-1) infection. *Neurology* (1991)
21. Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M.: FSL. *NeuroImage* **62**(9), 782–790 (2012)
22. Kim, H., Sabuncu, M.R.: Learning to compare longitudinal images. In: *Medical Imaging with Deep Learning*. pp. 3–13 (2023)
23. Lee, S.J., Park, K.W., Kim, L.S., Kim, H.: Association between frontal-executive dysfunction and speech-in-noise perception deficits in mild cognitive impairment. *Journal of Clinical Neurology* **14**, 513–522 (2018)
24. Liu, D., Lu, J., Wei, L., Yao, M., Yang, H., Lv, P., Wang, H., Zhu, Y., Zhu, Z., Zhang, X., Chen, J., Yang, Q.X., Zhang, B.: Olfactory deficit: a potential functional marker across the Alzheimer’s disease continuum. *Frontiers in Neuroscience* **18** (2024)
25. Milanini, B., Valcour, V.: Differentiating HIV-associated neurocognitive disorders from Alzheimer’s disease: an emerging issue in geriatric NeuroHIV. *Current HIV/AIDS Reports* **14**, 123–132 (2017)
26. Myznikov, A., Zheltyakova, M., Korotkov, A., Kireev, M., Masharipov, R., Dz., O., Jagmurov, Habel, U., Votinov, M.: Neuroanatomical correlates of social intelligence measured by the guilford test. *Brain Topography* **34**, 337–347 (2021)

27. Ouyang, J., Adeli, E., Pohl, K.M., Zhao, Q., Zaharchuk, G.: Representation disentanglement for multi-modal brain MRI analysis **12729**, 321–333 (2021)
28. Park, S.H., Zhang, Y., Kwon, D., Zhao, Qingyu, Z.N.M., Pfefferbaum, A., Sullivan, E.V., Pohl, K.M.: Alcohol use effects on adolescent brain development revealed by simultaneously removing confounding factors, identifying morphometric patterns, and classifying individuals. *Neurology* **8**, 8297 (2018)
29. Rohlfing, T., Maurer, C., Bluemke, D., Jacobs, M.: Volume-preserving nonrigid registration of mr breast images using free-form deformation with an incompressibility constraint. *IEEE Transactions on Medical Imaging*. **22**(6), 730–741 (2003)
30. Rohlfing, T., Zahr, N.M., Sullivan, E.V., Pfefferbaum, A.: The SRI24 multi-channel atlas of normal adult human brain structure. *Human Brain Mapping* **31**(5), 798–819 (2010)
31. Sanford, R., Cruz, A.L.F., Scott, S.C., Mayo, N.E., Fellows, L.K., Ances, B.M., Collins, D.L.: Regionally specific brain volumetric and cortical thickness changes in HIV-infected patients in the HAART era. *Journal of Acquired Immune Deficiency Syndromes* **74**, 563–570 (2017)
32. Sauty, B., Durrleman, S.: Progression models for imaging data with longitudinal variational auto encoders. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022. Lecture Notes in Computer Science*. **13431**, 3–13 (2022)
33. Son, G., Jahanshahi, A., Yoo, S.J., Boonstra, J.T., Hopkins, D.A., Steinbusch, H.W.M., Moon, C.: Olfactory neuropathology in Alzheimer’s disease: a sign of ongoing neurodegeneration. *BMB Reports* **54**, 295–304 (2021)
34. Tran, L., Yin, X., Liu, X.: Disentangled representation learning GAN for pose-invariant face recognition. In: *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. pp. 1415–1424 (2017)
35. Wang, X., Chen, H., Tang, S., Wu, Z., Zhu, W.: Disentangled representation learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* pp. 1–20 (2024)
36. Zahr, N.M., Pohl, K.M., Pfefferbaum, A., Sullivan, E.V.: Dissociable contributions of precuneus and cerebellum to subjective and objective neuropathy in HIV. *Journal of Neuroimmune Pharmacology* **14**, 436–447 (2019)
37. Zeghlache, R., Conze, P.H., El Habib Daho, M., Tadayoni, Ramin, M.P., Cochener, B., Quellec, G., Lamard, M.: Detection of diabetic retinopathy using longitudinal self-supervised learning. *Ophthalmic Medical Image Analysis* pp. 43–52 (2022)
38. Zhang, J., Zhao, Q., Adeli, E., Pfefferbaum, A., Sullivan, E.V., Paul, R., Valcour, V., Pohl, K.M.: Multi-label, multi-domain learning identifies compounding effects of HIV and cognitive impairment. *Medical Image Analysis* **75**, 102246 (2022)
39. Zhao, Q., Liu, Z., Adeli, E., Pohl, K.M.: Longitudinal self-supervised learning. *Medical Image Analysis* **71**, 102051 (2021)
40. Zhou, C., Yang, C., Ai, Y., Fang, X., Zhang, A., Wang, Y., Hu, H.: Valid olfactory impairment tests can help identify mild cognitive impairment: an updated meta-analysis. *Frontiers in Aging Neuroscience* **16**, 1349196 (2018)