UFFC



IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. XX, NO. XX, XXXX 2020

Disentangling Normal Aging from Severity of Disease via Weak Supervision on Longitudinal MRI

Jiahong Ouyang, Qingyu Zhao, Ehsan Adeli, Senior Member, IEEE, Greg Zaharchuk, and Kilian M. Pohl

Abstract—The continuous progression of neurological diseases are often categorized into conditions according to their severity. To relate the severity to changes in brain morphometry, there is a growing interest in replacing these categories with a continuous severity scale that longitudinal MRIs are mapped onto via deep learning algorithms. However, existing methods based on supervised learning require large numbers of samples and those that do not, such as self-supervised models, fail to clearly separate the disease effect from normal aging. Here, we propose to explicitly disentangle those two factors via weak-supervision. In other words, training is based on longitudinal MRIs being labelled either normal or diseased so that the training data can be augmented with samples from disease categories that are not of primary interest to the analysis. We do so by encouraging trajectories of controls to be fully encoded by the direction associated with brain aging. Furthermore, an orthogonal direction linked to disease severity captures the residual component from normal aging in the diseased cohort. Hence, the proposed method quantifies disease severity and its progression speed in individuals without knowing their condition. We apply the proposed method on data from the Alzheimer's Disease Neuroimaging Initiative (ADNI, N=632). We then show that the model properly disentangled normal aging from the severity of cognitive impairment by plotting the resulting disentangled factors of each subject and generating simulated MRIs for a given chronological age and condition. Moreover, our representation obtains higher balanced accuracy when used for two downstream classification tasks compared to other pre-training approaches. The code for our weak-supervised approach is available at https://github.com/ouyangjiahong/ longitudinal-direction-disentangle.

Index Terms—Weakly supervised learning, Disentanglement, Longitudinal MRI, Cognitive Impairment.

This research was in part funded by NIH grants MH113406 and AA005965. This study also benefited from Stanford Institute for Humancentered Artificial Intelligence (HAI) Google Cloud Credit.

J. Ouyang and Q. Zhao contributed equally.

J. Ouyang is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305.

Q. Zhao, E. Adeli, and K.M. Pohl are with the Department of Psychiatry and Behavioral Sciences, Stanford School of Medicine, Stanford, CA 94305.

E. Adeli is also with the Department of Computer Science, Stanford University, Stanford, CA 94305.

G. Zaharchuk is with the Department of Radiology, Stanford School of Medicine, Stanford, CA 94305.

K.M. Pohl is also with the Center for Health Sciences, SRI International, Menlo Park CA, 94025.

Corresponding author: K.M. Pohl (email: kilian.pohl@stanford.edu).

I. INTRODUCTION

1

Individuals suffering from neurological diseases (such as cognitive impairment) are often categorized into conditions (such as stable Mild Cognitive Impairment (sMCI), progressive MCI (pMCI), and Alzheimer's disease (AD) [1]). Accurately characterizing the effect of each condition on the human brain can contribute to a better understanding of the underlying neurological mechanisms and thus facilitate treatments in clinical settings. To do so, longitudinal MRI studies recruit normal controls and individuals of each condition, scan them repeatedly over time, and analyze their morphometric trajectories [2]–[4].

State-of-the-art analysis is often based on supervised learning of deep neural networks [4], [6]–[10] directly classifying longitudinal MRIs into normal controls and individuals associated with conditions of the disease (see Fig. 1(a)). However, supervised learning generally requires a large number of training samples for each condition, which are generally difficult to attain in longitudinal MRI studies [11]. Moreover, using exclusive labels to encode the classes ignores the underlying continuum across conditions regarding disease severity [9].

To reduce the need for large labelled training datasets, self-supervised learning models [12]-[20] derive a compact representation by mapping longitudinal MRIs into a latent space using label-independent information [5], [21]. One can then combine these self-supervised approaches with factor disentanglement [22]-[24] to extract directions in the latent space related to interpretable semantic information about the data. Common approaches for disentanglement (e.g., β -VAE [22], β -TCVAE [23], FactorVAE [24]) were created for crosssectional data, so the disentangled factors generally do not relate to time-related information (e.g., aging and disease progression). Existing works on time-series data, like audio and video [25]-[27], attempted to disentangle the time-invariant content (e.g., identity) from time-variant content (e.g., human body pose). However, these models were not designed to encode changes captured by longitudinal neuroimaging data, e.g., aging and disease progression. To overcome this limitation, we proposed Longitudinal Self-Supervised Learning (LSSL) [5] (Fig. 1(c)), which uses the order of visits to 'supervise' the disentanglement of brain aging encoded as a linear direction in the latent space. Our next model (referred to as LNE [28]) then relaxed the constraint of a single linear direction by



Fig. 1. Latent space of fully-supervised methods, weakly-supervised methods (Ours), and self-supervised methods ((LSSL [5]). Arrows represent the change within a subject between consecutive visits. Cyan arrows are of normal controls while arrows in yellow, purple, and brown are of subjects diagnosed with a condition. The background color represents the label used in the fully and weakly supervised setting. Note, while weakly-supervised learning (b) trains on subjects of all conditions (represented by arrows with four different colors), it only uses binary weak label to distinguish controls (cyan background) from diseased subjects (pink backgrounds).

deriving a non-linear smooth trajectory field. However, due to the absence of label information, both LSSL and LNE confined the disentanglement to one time-varying factor. Thus, these models assume that the effects of normal aging and disease progression lie on the same continuum so that they are encoded by the same direction in the latent space.

2

However, recent studies have shown that morphological brain changes associated with AD are different from normal aging [29]. To unravel these differences within a continuous space, we propose here a weakly supervised approach (Fig. 1(b)) that categorizes subjects as either normal control or diseased (i.e., belonging to one of the conditions). The direction associated with brain age is disentangled with respect to the control cohort by encouraging their aging trajectories to be fully encoded by that direction, while the direction linked to disease severity captures the residual after extracting 'normal aging' from the trajectories of diseased subjects. As normal controls are impartial to the severity, severity stages between individuals can be inferred solely from subject-specific progression without knowledge of their condition. By doing so, our model (similar to self-supervised approaches) can train on samples from conditions not of interest (e.g., AD subjects can be used for training when the primary analysis is on sMCI and pMCI).

We evaluated our method on disentangling brain age from cognitive impairment based on 632 longitudinal T1-weighted MRIs (ranging from 2 to 6 scans per participant) collected by the Alzheimer's Disease Neuroimaging Initiative (ADNI). The data set contained 185 NC, 193 sMCIs, 135 pMCIs, and 119 AD subjects. The weak label divides subjects into being normal controls or cognitively impaired (which was the collection of sMCI, pMCI, and AD). After training the model, plotting the disentangled factors of each sample in the latent space and simulating MRIs for any given chronological age and condition illustrates that the model properly disentangled normal aging from the severity of cognitive impairment. Note, the estimation of disease severity and disease progression speed cannot be produced by existing self-supervised or weakly-supervised methods. Moreover, compared to these alternative pre-training methods, the latent representation pretrained by our model obtains higher balanced accuracy on two downstream tasks: distinguishing sMCI from pMCI, and predicting the presence of amyloid plaques on positron emission tomography (PET).

II. METHOD

We derive our approach by first defining the latent space of 'trajectory vectors' where each vector represents the changes between two T1-weighted MRIs of a longitudinal sequence from the same subject (Section II-A). We then formalize the problem statement for disentangling the vectors into the effects associated with disease severity versus brain age (Section II-B). Finally, we encode this model via a loss function (Section II-C) and describe in Section II-D how to visualize the effect associated with a disentangled direction by generating synthetic brain MRIs. Note, for simplicity, we now drop the 'T1-weighted' from MRI.

A. Deriving Trajectory Vectors

As shown in Fig. 2, our model consists of an autoencoding structure that maps each MRI (or visit) to a point in the latent space. In doing so, each subject is encoded in that space by a trajectory (light blue for normal controls and pink for diseased subjects) across multiple visits (≥ 2). To reduce the problem of having an insufficient number of training samples, we do not directly train the model on the trajectories but instead compute vectors from all possible visit pairs, where the first visit in each pair precedes the second one. Note, the training weighs each MRI pair equally so that the model relies more on subjects with more observations.

We formalize a similar construction of the latent space as in [5]. Let $\mathcal{X} := \{x^1, \ldots, x^N\}$ be the collection of all MRIs and \mathcal{S} be the set of subject-specific MRI pairs (x^r, x^s) with x^r being scanned before x^s . As shown in Fig. 2, the encoder F maps an MRI to an M-dimensional latent space via z := $F(x) \in \mathbb{R}^M$. An MRI pair (x^r, x^s) is then mapped to $(z^r :=$ $F(x^r), z^s := F(x^s))$. The normalized vector is formulated as $\Delta z^{(r,s)} := (z^s - z^r)/\Delta t^{(r,s)}$, where $\Delta t^{(r,s)}$ is the time interval between the two MRIs. For convenience, we denote $\Delta z^{(r,s)}$ as Δz , and $\Delta t^{(r,s)}$ as Δt from now on. From the latent representations, the decoder H then reconstructs the input



Fig. 2. Overview of our method: an encoder projects a pair of MRIs (x^r, x^s) into the latent space resulting in a trajectory vector with (light) blue for normal controls and pink (or red) for diseased subjects. We encourage the direction of this vector to be consistent with the age direction τ_a (blue) for normal controls and the residual of that direction to be consistent with the disease severity direction τ_d (red) for diseased individuals.

MRIs, i.e., $\tilde{x}^r := H(z^r)$, $\tilde{x}^s := H(z^s)$. The reconstruction loss is used to guarantee the latent representation encodes all morphological information in the MRIs, which can be formulated as:

$$L_{recon} := \mathbf{E}_{(x^r, x^s) \sim \mathcal{S}} \left(\| x^r - \tilde{x}^r \|_2^2 + \| x^s - \tilde{x}^s \|_2^2 \right).$$
(1)

where $\mathbf{E}_{(x^r, x^s) \sim S}$ denotes the expected value with respect to all MRI pairs S and $\|\cdot\|_2$ represents the Euclidean norm.

B. Problem Statement

Our goal is to find two directions in the latent space, such that moving along one direction corresponds to normal aging (i.e., expected change induced by getting older) while moving along the other relates to disease progression. The remaining M-2 dimensions then encode other time-independent factors (e.g., sex, race) that are not of interest to this longitudinal study. To achieve such disentanglement, we first introduce the weak binary label y as being 0 for normal controls or 1 for diseased subjects. Note, a disease (such as cognitive impairment) can encompass several conditions (e.g., sMCI, pMCI, or AD), where each condition is associated with a certain level of disease severity. To make this problem tractable, we make the following relaxing assumptions:

Completeness. In line with many longitudinal studies of neurological diseases [30], [31], we assume that (brain) age and the severity of the disease are the only time-dependent factors. Thus, intra-subject changes captured by longitudinal brain MRIs can be fully explained by normal aging and disease progression, with disease progression only relevant to diseased adults (such as in the analysis of AD by ADNI [32], [33]).

Independence. We assume aging and disease progression are two independent factors, so that disease progression can be explicitly disentangled from the normal aging effect by modeling the changes along two orthogonal directions in the latent space. In the case of AD, this assumption is supported by prior studies that AD can not be simply regarded as accelerated aging [29], [34].

Homogeneity. While the two cohorts may not share the same distribution in age, we assume that their distribution over

the speed of normal aging is the same. This assumption is needed as some conditions, such as AD, are associated with accelerated aging [29]. Thus, we consider aging speeds faster than normal as part of the disease as also done in [30], [35].

As shown in Fig. 2, we denote the brain age direction as τ_a and the disease severity direction as τ_d . These two directions are designed to be unit vectors and to be strictly orthogonal (inline with independence assumption), so that the high-dimensional representations can be projected into a 2D space spanned by these two orthogonal directions (Fig. 2). Given the latent representation z of an MRI, we define the projection value of z on the disease severity direction as the estimated *disease severity* associated with that MRI, i.e., $\varphi_d := z^T \tau_d$. The projection on the age direction, i.e., $\varphi_a := z^T \tau_a$, is the estimated **brain age**. Now let Δz be associated with an MRI pair with $\Delta \varphi_a$ being the estimated aging speed between visits, then the component on the age direction can be computed by projecting Δz on τ_a , i.e., $\Delta z_a := (\Delta z^T \tau_a) \tau_a = \Delta \varphi_a \tau_a$. Similarly, let $\Delta \varphi_d$ be the estimated progression speed capturing the changes in disease severity between visits, then the component on the disease severity direction is $\Delta z_d := (\Delta z^T \tau_d) \tau_d = \Delta \varphi_d \tau_d$.

The completeness assumption suggests that $\Delta z = \Delta z_a + \Delta z_d$, i.e., the longitudinal changes are the combination between normal aging and the disease progression. For normal controls ideally we have $\Delta z = \Delta z_a$. An exact disentanglement would thus satisfy

$$\Delta z - \Delta z_a - y \cdot \Delta z_d = 0 \tag{2}$$

for each subject, while the distribution of aging speed $\Delta \varphi_a$ is the same for the two cohorts (according to the homogeneity assumption).

C. Loss Functions

We now design an objective function, whose minimum achieves the condition of disentanglement (i.e., Eq. (2)). Simply defining this function by the L2-norm over the left-hand side of Eq. (2) will lead to a non-informative solution, which is shrinking the magnitude of z towards zero. To avoid this

IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. XX, NO. XX, XXXX 2020

scenario, we now derive a cost function that instead minimizes the cosine loss between vectors [5].

4

For normal controls, τ_a should be parallel to the trajectory vector Δz . To impose this constraint in the model, we define the first term of the cost function to be a cosine loss favoring zero-angle between the two vectors, i.e.,

$$L_{da} := \mathbf{E}_{(x^r, x^s) \sim \mathcal{S}, y=0} \left(-\cos(\theta_{\langle \Delta z, \tau_a \rangle}) \right).$$
(3)

Note, S, y = 0 denotes the MRI pairs of normal controls so that, unlike in our prior work [5], the expected value is only over the normal controls. In [5], the constraint applied to both cohorts, which violates our homogeneity assumption as the function does not distinguish between normal and accelerated aging.

Moreover, we add a loss term based on the Kullback–Leibler (KL) divergence to explicitly impose the homogeneity assumption of having 'identical distribution of aging speed' for the two cohorts. More specifically, assuming the distribution of the aging speed is captured by the normal distribution $\mathcal{N}(\mu_0, \sigma_0)$ for normal controls and $\mathcal{N}(\mu_1, \sigma_1)$ for the diseased cohort, then the KL loss is defined as:

$$L_{kl} := \log(\frac{\sigma_1}{\sigma_0}) + (\frac{\sigma_0^2 + (\mu_0 - \mu_1)^2}{2\sigma_1^2}) - \frac{1}{2}.$$
 (4)

To enforce the completeness assumption, we propose a penalty loss term based on the ratio of the norm of $\Delta \varphi_d$ between the normal controls and diseased cohort:

$$L_{pen} := \frac{\mathbf{E}_{y=0} \|\Delta \varphi_d\|_2}{\mathbf{E}_{y=1} \|\Delta \varphi_d\|_2}.$$
(5)

Note, the minimum of that ratio is achieved when $\mathbf{E}_{y=0} \|\Delta \varphi_d\|_2 = 0$, i.e., normal controls have no disease component.

Finally, for diseased subjects, the residual of the age direction (i.e., $\Delta z - \Delta z_a$) should be parallel to the disease severity direction τ_d . We model this assumption via

$$L_{dd} := \mathbf{E}_{(x^r, x^s) \sim \mathcal{S}, y=1} \left(-\cos(\theta_{\langle \Delta z - \Delta z_a, \tau_d \rangle}) \right).$$
(6)

In doing so, the proposed method can be applied to scenarios where diseases can be either relevant or irrelevant to brain age.

The complete objective function, whose minimum fulfills the requirement of Eq. (2), is then the weighted combination of prior loss functions, i.e.,

$$L := \lambda_{recon} L_{recon} + \lambda_a L_{da} + \lambda_d L_{dd} + \lambda_{pen} L_{pen} + L_{kl},$$
(7)

where λ_{recon} , λ_a , λ_d , and λ_{pen} are hyperparameters that balance the losses.

D. Visualizing the Effect of Normal Aging and Disease

With the disentangled directions, we can visualize the effect of normal aging by reconstructing the average brain MRI of the control cohort, whose z varies along the age direction. The disease effect can be visualized by varying z along the disease severity direction to reconstruct diseased MRIs. Let $\{z^i | i = 1, ..., N^C\}$ be the latent representations of N^C normal controls, we generate a simulated MRI of controls \hat{x}^C at a given brain age $\hat{\varphi}_a$ by reconstructing the following latent representation

$$\hat{z}^{C} := \hat{\varphi}_{a} \tau_{a} + \frac{1}{N^{C}} \sum_{i=1}^{N^{C}} [z^{i} - \varphi_{a}^{i} \tau_{a}],$$
(8)

where the first term corresponds to the component along the age direction and the second term captures the factors independent from brain age, i.e., the group average of the components orthogonal to τ_a .

Assuming the disease consists of K conditions of various severity levels, the simulated brain MRI at age $\hat{\varphi}_a$ for the k^{th} condition (with N^{D_k} subjects) can be reconstructed from:

$$\hat{z}^{D_k} = \hat{\varphi}_a \tau_a + \left(\frac{1}{N^{D_k}} \sum_{i=1}^{N^{D_k}} \varphi_d^i\right) \tau_d + \frac{1}{N^{D_k}} \sum_{i=1}^{N^{D_k}} \left(z^i - \varphi_a^i \tau_a - \varphi_d^i \tau_d\right),$$
(9)

where the first term is the aging component, the second term corresponds to disease severity, and the last term captures the factors independent from brain age and disease.

III. EXPERIMENTAL SETUP

A. Data

We evaluated our method on 632 subjects' longitudinal T1w MRIs (consisting of 2389 individual MRIs that successfully were pre-processed) from ADNI-1. Each subject had at least two (and up to 6) T1w MRI scans (Table I)), which were acquired via a 1.5T 3D MPRAGE sequence defined across GE, Siemens, and Phillips scanners (TR/TE = 2300-3000/3-4 ms; flip angle = $8-9^{\circ}$; section thickness = 1.2 mm; 256 reconstructed axial sections) [36]. According to ADNI, electronic protocols were supplied to each MRI scanner vendor to minimize inconsistencies expected to arise from building the protocol manually on individual scanners and the imaging protocol remained the same for each individual during the study period of ADNI-1 [37].

The data set consists of 185 normal controls (age: 75.57 ± 5.06 years), 119 subjects with AD (age: 75.17 ± 7.57 years), 135 subjects diagnosed with pMCI (age: 75.91 ± 5.35 years), and 193 subjects diagnosed with sMCI (age: 75.63 ± 6.62 years). pMCI is defined as those MCI subjects that progressed into AD within 3 years after baseline, whereas sMCI did not [38]. There was no significant age difference among sMCI, pMCI and AD individuals (p >0.3, two-sample *t*-test).

We assigned a binary weak label to each individual with y = 0 for normal controls and y = 1 for cognitively impaired individuals, which include sMCI, pMCI, or AD subjects. There was no significant age difference (p=0.62; two-sample *t*-test) between controls (age: 75.57 ± 5.06 years) and the cognitive impaired group (75.59 ± 6.42 years).

In line with our prior studies [5], [10], [28], all longitudinal MRIs were preprocessed by a pipeline composed of denoising, bias field correction, skull striping, affine registration to a template, re-scaling to a $64 \times 64 \times 64$ volume, and transforming image intensities to z-scores. At the sacrifice of image

Transactions on Medical Imaging JIAHONG OUYANG *et al.*: PREPARATION OF PAPERS FOR IEEE TRANSACTIONS ON MEDICAL IMAGING



Fig. 3. Network structure of the proposed method. Blue blocks correspond to the encoder that reduces an input MRI to a 1024 dimensional latent representation z, from which the decoder (green) reconstructs the MRI. τ_a and τ_d are constructed as the output of a dummy network. The last dimension of τ_d enforces the orthogonality constraint $\tau_a \perp \tau_d$. The relationship between variables is further illustrated in Fig. 2.

 TABLE I

 NUMBER OF VISITS FOR SUBJECTS OF FOUR DIAGNOSIS COHORTS.

Diagnosis	2 visits	3 visits	4 visits	5 visits	6 visits
NC	28	32	43	60	22
sMCI	44	49	44	46	10
pMCI	17	31	28	45	14
AD	27	30	61	1	0

resolution, the downsampling enables the design of a compact encoder model with a relatively small number of network parameters, which can effectively boost training speed. Finally, we constructed 2668 MRI pairs by the criteria that each pair was from the same subject and scans were at least one year apart (i.e., contain sufficient morphological changes).

B. Implementation Details

We augmented the training set as we did in [28], i.e., by applying the same random shift (within 4 pixels), rotation (within 2 degrees) and random flipping of brain hemispheres on each pair of MRIs. By doing so, we preserve the intra-subject changes that our model aims to learn (i.e., aging and disease effects). This augmentation strategy also allowed for direct comparison with our previous works [5], [28]. Regarding the architecture, our model was based on an Encoder-Decoder structure [39]. Specifically, let EB_k denote an Encoder Block, i.e. a stack of Convolution layer (k channels, kernel size of $3 \times 3 \times 3$) followed by a BatchNorm, LeakyReLU (with slope of 0.2) and a MaxPool layer (kernel size of 2), and DB_k as Decoder Block, i.e., a stack of Convolution layer (k channels, kernel size of $3 \times 3 \times 3$) followed by a BatchNorm, LeakyReLU (with slope of 0.2) and a MaxPool layer (kernel size of 2). Then the architecture of our model was a EB_{16} -EB₃₂-EB₆₄-EB₁₆-DB₆₄-DB₃₂-DB₁₆-DB₁₆ followed by a convolution layer for the final reconstruction. As shown in Fig. 3, the encoder resulted in a 1024-dimensional representation space. We estimate τ_a and τ_d by regarding them as the output of two dummy layers. Specifically, a dummy constant scalar of 1 was fed to a dense layer (top orange block) producing the 1024 dimensional vector τ_a , where we denote the first 1023 dimensions as τ'_a and the last dimension as τ''_a . The constant was also fed to a separate dense layer (bottom orange block) producing another 1023 dimensional vector τ'_d , which was then concatenated with $-\frac{\tau'_a \tau' \tau'_d}{\tau''_a}$ to produce the final 1024 dimensional τ_d . By design, τ_d was orthogonal to τ_a , both of which were then turned into unit vectors via normalization.

5

Due to the complexity of the objective function, and inspired by the idea of curriculum learning [40], [41], we designed a coarse-to-fine step-wise training strategy that gradually added loss terms to the objective function. On all training subjects, we first optimized over $\lambda_{recon}L_{recon} + \lambda_a L_{da}$ so that disease progression was modelled as an accelerated aging effect. We then added the loss function associated with the disease severity direction (i.e., $\lambda_{recon}L_{recon} + \lambda_a L_{da} + \lambda_d L_{dd}$) to disentangle the disease progression from normal aging. Finally, the disentanglement was further improved by minimizing the entire object function as specified in Eq. (7) In line with LSSL [5], we set λ_{recon} and λ_a to be equal as the number of MRIs (N=2389) is about the same as the number of MRI pairs (N=2668). We then set the weights for the remaining regularization terms to be half of λ_{recon} so that the loss values of all terms are of equal scale. The networks were trained for 25 epochs by the Adam optimizer [42] with learning rate of 5×10^{-4} and weight decay of 10^{-5} for each of the three phases.

C. Evaluation

The approach was evaluated in two steps. We first ensured that brain aging was accurately disentangled from disease severity. Next, we assessed the use of this disentanglement for downstream tasks by comparing its accuracy to those of alternative approaches on several clinically-relevant tasks.

Quality of disentanglement in the latent space: A successful disentanglement would result in the direction τ_d stratifying the severity level of cognitive impairment within the diseased individuals. To assess the quality of the stratification, we first trained the model using the weak labels of all subjects and then examined whether the resulting representations could differentiate sMCI, pMCI, and AD within the diseased cohort. To do so, we selected trajectory vectors Δz associated with adjacent visits of each subject and examined the differences

IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. XX, NO. XX, XXXX 2020

in $\Delta \varphi_a$ (aging speed) and $\Delta \varphi_d$ (disease progression speed) across cohorts by two-tailed two-sample t-tests (p < 0.05was viewed as significant). We followed up this quantitative analysis by projecting the trajectory vector Δz as a dot in the 2D space spanned by the age direction τ_a and disease severity direction τ_d . We color-coded the dots by cohorts to visualize their differences in the Δz_d component.

6

In a second experiment, we inspected the correlation between the estimated progression speed and the changing speed of Alzheimer's Disease Assessment Scale–Cognitive Subscale (ADAS-Cog). ADAS-Cog quantifies the severity of cognitive impairment by having subjects complete 11 tasks that assess the cognitive domains of memory, language, and praxis. The outcome of this assessment is a single value ranging from 0 to 70 with higher scores suggesting worse cognition [43]. In line with the computation of $\Delta \varphi_d$, we defined the changing speed of ADAS-cog as the difference in ADAS-cog between the two visits divided by the time interval. Finally, the above two experiments were repeated on φ_d ; i.e., we tested the difference in φ_d across the three conditions and correlated φ_d with ADAS-Cog.

Evaluating Quality of Representations for Two Downstream Tasks: Distinguishing MCI subjects who will eventually develop AD (pMCI) from stable MCI subjects (sMCI) within a short time frame (3 years) is helpful to identify patients who may need more medical care or may be appropriate to include in clinical trials of AD disease modifying therapies. To highlight the potential clinical value of the proposed method, the first downstream task was on **sMCI vs. pMCI classification**.

For the second downstream task, we used the learned representations to predict brain amyloid status (i.e., positive vs. negative status) [44]. The presence of brain β -amyloid $(A\beta)$ plaques is a defining feature of Alzheimer's disease. Traditionally, PET imaging using A β -specific radiotracers (such as 18-florbetapir) is used to quantify brain $A\beta$ plaque load. However, PET imaging is expensive and exposes the subjects to radiation. Moreover, PET is not available in majority of the hospitals around the world. Thus, predicting this PET biomarker from MRIs is significant for AD diagnosis. As MRI and PET acquisition were usually not acquired in the same visit by ADNI and subjects visits were generally 6-months apart, we considered the amyloid status derived from the 18F-AV45 PET acquired within 3 months from an MRI visit as the 'ground truth' of that visit. As suggested by prior longitudinal studies [45]–[47], we used the standardized uptake value ratio (SUVR) for each PET scan based on a composite reference region consisting of cerebellum, brainstem/pons, and eroded cortical white matter (described in 'UC Berkeley - AV45 Analysis Methods' which is available on ADNI [48]). Following the definition in [49], [50], an SUVR above 0.79 was labelled positive and otherwise the scan was labelled as negative.

Considering our method as a pre-training scheme that provides a representation that can potentially benefit downstream supervised tasks, we pre-trained the encoder based on our disentanglement approach on all subjects. Afterward, we assessed the accuracy of the resulting representation with respect to each of the two downstream tasks via five-fold

TABLE II

PROPERTIES OF PRE-TRAINING METHODS FOR LEARNING REPRESENTATIONS FOR DOWNSTREAM CLASSIFICATION. 'S' DENOTES SELF-SUPERVISED, AND 'W' DENOTES WEAKLY-SUPERVISED.

		Using	
	Weakly/Self-	Longitudinal	
	Supervised	Pair	Disentanglement
AE [54]	S	×	X
VAE [39]	S	×	X
SimCLR [55]	S	\checkmark	X
LSSL [5]	S	\checkmark	Brain age
LNE [28]	S	\checkmark	Brain age
wMLP [51]	W	×	X
C-VAE [56]	W	×	X
wLSSL1	W	\checkmark	Brain age
wLSSL2	W	\checkmark	Brain age
			Brain age
Ours	W	\checkmark	Disease Severity

cross-validation. For each run, one fold was used for testing, while the remaining folds were utilized for training, among which 10% subjects were held out for validation. The 5 folds were split based on subjects so that all MRIs of a subject were assigned to the same fold. For sMCI vs. pMCI classification, we tested on two types of input features: 1) a visitlevel prediction used z of each visit; and an MRI-pair-level prediction concatenated z of the first visit with Δz between visits, denoted as $z \& \Delta z$. The prediction target was always the sMCI vs. pMCI label of the subject. Using each feature type, we trained the classifier in two settings. We first performed the classification directly on the features (frozen) using a multi-layer perceptron (MLP) [51] with two dense layers of dimension 1024 and 64 with LeakyReLU activation [52]. In the second setting, we fine-tuned the features by incorporating the encoder F before the MLP classifier and trained the entire model in an end-to-end manner. Classification accuracy was measured by balanced accuracy (BACC) [53] accounting for different number of training samples in each group. We also conducted an ablation study to examine the impact of omitting the two regularization terms (Eq. 4, 5) on the BACC. Lastly, the cross-validation was repeated for PET amyloid status classification. Since the status was defined per visit, the prediction target was the status at the second visit when using $z \& \Delta z$ as input features.

We compared the BACC to models using the same architecture with encoders pre-trained by other representation learning methods. To the best of our knowledge, there exists no longitudinal approach that can disentangle aging and disease effects within a latent space. We therefore compared to unsupervised (AE [54], VAE [39]) and self-supervised (SimCLR [55], LSSL [5], LNE [28]) methods. Specifically, we adapted SimCLR to our longitudinal setting by treating two MRIs of the same subject (with the same shift and rotation augmentation) as a positive pair for self-supervised training [55]. Next, we also implemented several weakly supervised methods (see first column of Table II) for pre-training the encoder. First, wMLP concatenated the encoder with an MLP classifier [51] to predict the binary weak label, while a C-VAE JIAHONG OUYANG et al.: PREPARATION OF PAPERS FOR IEEE TRANSACTIONS ON MEDICAL IMAGING



Fig. 4. (a) Distribution of aging speed $\Delta \varphi_a$ for control and diseased subjects; (b) Distribution of progression speed $\Delta \varphi_d$; (c) Distribution of $\Delta \varphi_d$ with respect to the three conditions; (d) Each dot represents the aging and progression speed between a pair of MRIs from adjacent visits. The color of dots represents cohort assignment. In all sub-figures, black dashed lines denote zero speed. Note, we omitted from the axes their scale as the absolute value of aging speed and progression speed has no practical meaning.

[56] learned the latent representation of each MRI conditioned on its binary label. We also expanded the LSSL approach to work with weak labels by implementing two variants: similar to wMLP, wLSSL1 consisted of the encoder followed by a classifier in order to estimate the weak binary label; wLSSL2 estimated two label-specific directions in the latent space by using two cosine losses, one computed with respect to the controls and the other with respect to the diseased cohort. Table II summarizes the properties of all methods used in our comparison.

IV. RESULTS

We now review the findings of the experiments in the order they were described in the prior section.

A. Plotting Disentangled Values

As the proposed method performed disentanglement by focusing on the differences within pairs of MRIs, we first review our findings with respect to the aging speed $\Delta \varphi_a$ and the progression speed $\Delta \varphi_d$ (Fig. 4) and then complement those findings by analyzing age φ_a and disease severity φ_d (Fig. 6).

Fig. 4(a) shows no significant difference in aging speed $\Delta \varphi_a$ between the control (NC) and the cognitive impairment (CI) cohort (p=0.25). This indicates the KL loss L_{kl} (Eq. (4)) effectively extracted a common pattern of normal aging across the two cohorts. That the normal aging was similar across cohorts is further supported by Fig. 5(a), which shows similar developmental patterns across cohorts. The pattern for each cohort was created by subtracting the synthetic MRI (see also Section II-D) at age 60 years (first column) from that at different ages. Note, the estimated brain age significantly correlated with the actual chronological age (p < 0.01, Fig. 6(a)) so that the chronological ages were related to the estimated brain ages φ_a based on a linear function that mapped the range (i.e., max and min values) of brain ages to the one of chronological ages.

Furthermore, normal controls had close to zero progression speed (shown by the horizontal dashed line in Fig. 4(b)+(d))

indicating the efficacy of the regularization L_{pen} . However, when investigating the progression speed of the three conditions (Fig. 4(c)+(d)), sMCI had significantly smaller $\Delta \varphi_d$ compared to pMCI and AD (p < 0.01). Beyond progression speed, the disease severity of sMCI subjects was lower than those with pMCI or AD (both with p < 0.01, Fig. 6(b)). These findings are visually confirmed by the differences in simulated MRIs of each cohort to that of the normal controls at a specific age (Fig. 5(b)). While for all conditions we observed enlarged ventricle and brain atrophy, those morphological changes were least pronounced for sMCI. Furthermore, sMCI converged with increasing age to the brain of normal controls, while brains from pMCI and AD had visually distinct yet less pronounced morphological changes compared to effects at younger ages. These results agree with recent findings that brain atrophy of early onset AD patients is distinctly different to age-matched controls, but less so when comparing older AD patients from older controls [57].

7

The progression speed estimated by our method also aligned well with the actual speed of cognition decline, as larger $\Delta \varphi_d$ corresponded to faster ADAS-cog decline (light yellow in Fig. 7(a)). The Pearson correlation (r=0.27, p < 0.01, Fig. 7(b)) between the two variables was achieved solely using T1 structural data without supervision on ADAS-cog. Interestingly, the regression line (in red) almost goes through the origin of that coordinate system indicating that when no progression in disease severity was detected by our model then this accurately reflected stability in the cognitive score. Beyond progression speed, disease severity itself was correlated with the ADAS-cog score (p < 0.001) (Fig.6(c)). Interestingly, the slope of the fitted lines was larger in AD than pMCI, which suggests the difference in ADAS-cog score between the two cohorts was more pronounced than the difference in the estimated disease severity. This is potentially due to the fact that cognitive impairment is not only associated with structural brain changes but also with functional ones [58]. Thus changes cannot be solely explained by structural biomarkers [59] and hence requires further multi-modal analysis.

In summary, all these findings suggest that the proposed

IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. XX, NO. XX, XXXX 2020

method not only properly disentangled brain age from disease severity (p>0.1 for Pearson's correlation) but also was able to learn the progression of cognitive impairment from the ordinal information of subject-specific MRI pairs without knowledge about ADAS-cog or the three conditions (sMCI, pMCI, and AD).

B. sMCI vs. pMCI classification

8

According to Table III, the representations learned by the proposed method achieved significantly more accurate predictions (p < 0.05, DeLong's test) than all baselines in 3 out of 4 scenarios. The one exception was using z and fine-tuning the encoder, in which case LNE [28] achieved the highest accuracy (69.6%) as it can explicitly model non-linear aging directions. Another observation is that the accuracy of the proposed method using z and Δz (a.k.a., z & Δz) with a frozen encoder was more accurate (70.9%) than any implementation just relying on z (including those with fine-tuned encoders). This indicates the progression speed of cognitive impairment was an important marker for distinguishing pMCI from sMCI. This would also explain why cross-sectional methods were generally outperformed by longitudinal ones as the latter explicitly modeled the morphological change between time points, which led to more accurate estimates of Δz than computed by cross-sectional models.

Compared with the model that directly classifies sMCI and pMCI individuals (i.e., 'No pretrain'), the proposed method pre-trained the encoder by augmenting the data set with the control and AD subjects. Unlike other unsupervised or weakly-supervised pre-training schemes (i.e., wMLP, C-VAE, wLSSL1 and wLSSL2), our model leveraged the weak label (without knowing the condition) to explicitly disentangle the disease factor from the aging factor, which is particularly beneficial for modeling stages of severity where each stage is represented by a limited number of training samples. This statement is supported by excluding the AD cohort in pretraining resulting in a significant drop in balanced accuracy from 70.9 to 70.1 (p < 0.05, DeLong's test) in the sMCI vs. pMCI classification (based on z & Δ z and a frozen encoder). This conclusion is further supported by reported accuracy score of 74.2% of our method that is higher than any other published accuracy on sMCI vs. pMCI classification just relying on structural MRI (Table IV).

Ablation study. As we have qualitatively shown the importance of L_{kl} (Eq. (4)) (ensures the homogeneity assumption of having same distribution of aging speed) and L_{pen} (Eq. 5) (enforces the completeness assumption that normal aging can fully explain the changes in MRI for the control cohort) for disentanglement in Section IV-A, we now quantitatively assess their impact on the BACC in the sMCI vs. pMCI classification task. As shown in Table V, using all loss terms during pre-training obtained the highest accuracy for three settings compared to omitting either regularization term, while omitting both terms achieved the worst performance. In the scenario of using z as input features and fine-tuning the encoder, omitting L_{pen} resulted in slightly higher accuracy (68.5%) than including it (68.4%), which was much lower than

TABLE III

Average balanced accuracy for sMCI vs. pMCI classification with and W/O fine-tuning the encoder using representation zand representation $z \& \Delta z$. The best result in each column is in bold while the second best is underlined. '*' denotes weakly-supervised, and 't' denotes longitudinal.

	sMCI vs. pMCI			
Methods	z		$z \ \& \ \Delta z$	
	Frozen	Fine-tune	Frozen	Fine-tune
No pretrain	-	64.2	-	69.3
AE [54]	60.9	65.3	62.6	69.5
VAE [39]	60.8	63.4	61.3	63.8
wMLP * [51]	60.3	66.2	62.1	69.6
C-VAE * [56]	60.6	64.2	61.6	64.1
SimCLR [†] [55]	<u>61.4</u>	66.7	63.3	69.5
LSSL † [5]	59.3	68.2	69.4	71.2
LNE [†] [28]	59.8	69.6	<u>70.6</u>	<u>73.4</u>
wLSSL1 *†	<u>61.4</u>	<u>68.6</u>	70.2	72.1
wLSSL2 *†	60.8	68.4	<u>70.6</u>	72.8
Ours *†	62.2	68.4	70.9	74.2

TABLE IV

COMPARISON OF THE PROPOSED METHOD WITH OTHER TRADITIONAL METHODS AND DEEP-LEARNING-BASED METHODS IN SMCI VS. PMCI CLASSIFICATION ON ADNI DATASET. THE PROPOSED METHOD ACHIEVED BEST BACC AMONG ALL METHODS THAT WERE SOLELY BASED ON STRUCTURAL MRI.

Method	Modalities	sMCI vs. pMCI	BACC
Cross-sectional			
Liu et al. [60]	MRI	465/205	62.2
Zu et al. [61]	MRI, PET	56/43	69.0
Suk et al. [62]	MRI	128/76	63.8
Lin et al. [63]	MRI	100/164	73.0
Huang et al. [64]	MRI, PET	441/326	76.9
Zhou et al. [65]	MRI, PET, SNP	205/157	74.3
Zhou et al. [66]	MRI, PET	114/71	78.3
Longitudinal			
Gray et al. [67]	MRI, PET	64/53	62.7
Cui et al. [4]	MRI	236/167	71.7
Platero et al. [68]	MRI, tests	215/206	77.1
Ouyang et al. [28]	MRI	193/135	73.4
Ours	MRI	193/135	74.2

the overall highest accuracy of our implementation (74.2%). This demonstrates the necessity of each proposed component in learning informative representations stratifying disease severity across conditions.

C. Amyloid status classification

As shown in Table VI, the proposed method yielded the best accuracy of classifying amyloid status in all four scenarios, suggesting the superiority of the representations learned by our method. By disentangling the disease progression from normal aging, representation z captured more information about the status of AD, which was correlated with amyloid status. Unlike for sMCI vs. pMCI classification, including Δz in the classifier did not greatly boost accuracy, which indicates that the amyloid status might be less related to the progression speed but rather the severity of cognitive impairment.

V. DISCUSSION

9

Transactions on Medical Imaging JIAHONG OUYANG et al.: PREPARATION OF PAPERS FOR IEEE TRANSACTIONS ON MEDICAL IMAGING



Fig. 5. (a) The intensity difference between the simulated brain at a categorical age and the one at age 60 years (first column). (b) For each condition (i.e., row), the corresponding column shows the difference between its simulated brain and that of the normal control cohort (first row) at the corresponding categorical age.



Fig. 6. (a) Correlation between the estimated brain age and the actual chronological age. Each blue line connects adjacent visits. The red line corresponds to the group-level relation fitted by a robust linear mixed effect model. (b) Distribution of disease severity by condition. (c) Correlation between disease severity and the ADAS-cog score colored according to condition. The thick straight line associated with each condition represents the average trajectory fitted by a robust linear mixed effect model.



Fig. 7. In (a), each dot represents the aging and progression speed between a pair of MRIs from adjacent visits. The color of each dot indicates the speed of the ADAS-cog score (purple = slow, yellow = fast). (b) illustrates the correlation between estimated progression speed and ADAS-cog speed. In all sub-figures, black dashed lines correspond to zero. Note, we omitted from the axes their scale as the absolute value of aging speed and progression speed has no practical meaning.

0278-0062 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information. Authorized licensed use limited to: University of Southern California. Downloaded on May 10,2022 at 23:00:27 UTC from IEEE Xplore. Restrictions apply.

TABLE V ABLATION STUDY ON SMCI VS. PMCI CLASSIFICATION. THE BEST RESULT IN EACH COLUMN IS IN BOLD.

	sMCI vs. pMCI			
Methods	z		$z \ \& \ \Delta z$	
	Frozen	Fine-tune	Frozen	Fine-tune
No L_{kl}, L_{pen}	59.2	68.2	69.1	71.4
No L_{kl}	59.8	68.0	69.8	72.6
No L_{pen}	59.4	68.5	69.3	71.4
All terms	62.2	68.4	70.9	74.2

TABLE VI

Average balanced accuracy for amyloid positive vs. Negative classification with and W/O fine-tuning the encoder using representation z and representation $z \& \Delta z$. The best result in each column is in bold while the second best is underlined. '*' denotes weakly-supervised, and '†' denotes longitudinal.

Methods	Amyloid positive vs. negative			
Wiethous	z		$z \ \& \ \Delta z$	
	Frozen	Fine-tune	Frozen	Fine-tune
No pretrain	-	78.2	-	78.4
AE [54]	60.2	77.7	60.6	78.2
VAE [39]	60.4	77.9	60.3	78.2
wMLP * [51]	60.7	78.1	61.2	79.0
C-VAE * [56]	60.1	77.2	60.4	78.0
SimCLR [†] [55]	60.4	78.6	61.0	77.9
LSSL † [5]	60.8	79.0	61.5	79.0
LNE [†] [28]	60.6	78.8	<u>61.6</u>	<u>79.4</u>
wLSSL1 *†	<u>61.0</u>	79.4	<u>61.6</u>	78.9
wLSSL2 *†	60.6	78.7	61.3	78.6
Ours	61.2	79.4	62.2	80.1

The advantage of the proposed method is that it explicitly disentangles the disease progression from normal aging. Thus, our model provides a unique way to understand the disease status (severity and progression) by directly learning from the longitudinal MRIs rather than using human-defined metrics related to the disease. More specifically, our proposed method captures a disease via φ_d representing disease severity and $\Delta \varphi_d$ encoding the progression speed of a disease. Moreover, it yields an intuitive way for interpreting and visualizing the representation, i.e., projecting the high dimensional representation into 2D-space supported by the aging and disease directions, as shown in Fig. 5(d) and Fig. 7(a).

Beyond those advantages, the accuracy of our method could be negatively impacted by the completeness assumption, i.e., that any longitudinal changes are only caused by aging and the disease (see Section II). In practice, longitudinal changes are also caused by other time-dependent factors, such as diseases whose risk increases with age [69] and AD comorbidities [70]– [73]. Those factors are only implicitly accounted for as they are either modeled as part of aging (if they are independent of AD), disease severity (if they related to AD), or captured by the remaining vectors (those that are orthogonal to both aging and disease severity).

Second, confounding effects of time-invariant factors (such as sex) were not considered during training. We studied the effect of sex via a post-hoc analysis, which revealed that the male cohort had a significant higher brain age φ_a and

disease severity φ_d (both p < 0.01, two-sample *t*-test) than the female one. These findings align with the demographics of the ADNI data set as the male cohort was significantly older and contained a significantly higher percentage of cognitive impaired subjects than the female cohort (both p < 0.01). However, the sex difference was not significant for aging speed $\Delta \varphi_a$ (p=0.1) and progression speed $\Delta \varphi_d$ (p=0.07).

Lastly, in our experiments, the proposed method was pretrained on the whole dataset, but it has the potential to be embedded into a cross-validation setting. To prove this concept, we also applied the pre-trained method on a separate hold-out dataset consisting of 200 MRI pairs of 200 subjects (50 subjects for each cohort) from the ADNI2/3 data set (i.e., no overlapping with the pre-trained dataset). The cohorts in the hold-out set were age matched to the pre-trained dataset (p > 0.05, t-test). As on pre-trained data set (Fig. 4 (b)(c)), the progression speed measured on the hold-out set of pMCI or AD subjects was significantly faster than NC or sMCI (p < 0.01 for all four comparisons including NC vs. pMCI, NC vs. AD, sMCI vs. pMCI, and sMCI vs. AD). Furthermore, the BACC for the sMCI vs. pMCI classification for the frozen encoder with respect to z was 61.4 %, which is insignificantly different to the 62.2 % reported in Table III (p > 0.1, Fisher's exact test). Insignificantly different was also the difference in BACC with respect to the frozen encoder applied to $z\&\Delta z$, for which we recorded 70.3% (versus 70.9 %, p > 0.1, Fisher's exact test) on the hold-out dataset. Moreover, the model with the fine-tuned encoder also had a balanced accuracy of 68.2 % (vs. 68.4 %) using z and 73.5% (vs. 74.2 %) using $z\&\Delta z$ (both have p > 0.1, Fisher's exact test). In other words, our derived encoder can generalize to unseen data.

VI. CONCLUSION

In this work, we proposed a representation learning framework based on weakly-supervised learning that takes advantage of the repeated MRI scans acquired by longitudinal studies to explicitly disentangle normal brain aging and disease progression. When applied to analyze the longitudinal T1weighted MRIs from ADNI, the proposed method not only accurately quantified the severity level and the dynamic progression speed of cognitive impairment associated with sMCI, pMCI, and AD, but also visualized for each condition the effects of the disease on brain morphometry. When used as a pretraining strategy, our method resulted in latent representations with superior accuracy in differentiating sMCI from pMCI and positive from negative amyloid status compared with the ones learned by other unsupervised and self-supervised methods.

REFERENCES

- R. Bajo *et al.*, "Differential patterns of connectivity in progressive mild cognitive impairment," *Brain Connectivity*, vol. 2, no. 1, pp. 21–24, 2012.
- [2] J. L. Bernal-Rusiel, D. N. Greve, M. Reuter, B. Fischl, and M. R. Sabuncu, "Statistical analysis of longitudinal neuroimage data with linear mixed effects models," *NeuroImage*, vol. 66, pp. 249–260, 2013.
- [3] M. Bilgel, J. L. Prince, D. F. Wong, S. M. Resnick, and B. M. Jedynak, "A multivariate nonlinear mixed effects model for longitudinal image analysis: Application to amyloid imaging," *NeuroImage*, vol. 134, pp. 658–670, 2016.

Transactions on Medical Imaging JIAHONG OUYANG et al.: PREPARATION OF PAPERS FOR IEEE TRANSACTIONS ON MEDICAL IMAGING

- [4] R. Cui and M. Liu, "RNN-based longitudinal analysis for diagnosis of Alzheimer's disease," *Computerized Medical Imaging and Graphics*, vol. 73, pp. 1–10, 2019.
- [5] Q. Zhao, Z. Liu, E. Adeli, and K. M. Pohl, "Longitudinal self-supervised learning," *Medical Image Analysis*, vol. 71, p. 102051, 2021.
- [6] Z. C. Lipton, D. C. Kale, C. Elkan, and R. Wetzel, "Learning to diagnose with LSTM recurrent neural networks," arXiv preprint arXiv:1511.03677, 2015.
- [7] R. Santeramo, S. Withey, and G. Montana, "Longitudinal detection of radiological abnormalities with time-modulated LSTM," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Lecture Notes in Computer Science*, vol. 11045, pp. 326–333, 2018.
- [8] R. Gao et al., "Distanced LSTM: Time-distanced gates in long shortterm memory models for lung cancer detection," in *International Workshop on Machine Learning in Medical Imaging, Lecture Notes in Computer Science*, vol. 11861, pp. 310–318, 2019.
- [9] M. M. Ghazi et al., "Training recurrent neural networks robust to incomplete data: Application to Alzheimer's disease progression modeling," *Medical Image Analysis*, vol. 53, pp. 39–46, 2019.
- [10] J. Ouyang *et al.*, "Longitudinal pooling & consistency regularization to model disease progression from MRIs," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 6, pp. 2082–2092, 2020.
- [11] A. Carass *et al.*, "Longitudinal multiple sclerosis lesion segmentation: resource and challenge," *NeuroImage*, vol. 148, pp. 77–102, 2017.
- [12] S. Gidaris, P. Singh, and N. Komodakis, "Unsupervised representation learning by predicting image rotations," in *International Conference on Learning Representations*, 2018.
- [13] R. Zhang, P. Isola, and A. A. Efros, "Colorful image colorization," in European Conference on Computer Vision, pp. 649–666, 2016.
- [14] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *European Conference on Computer Vision*, pp. 184–199, 2014.
- [15] M. Noroozi and P. Favaro, "Unsupervised learning of visual representations by solving jigsaw puzzles," in *European Conference on Computer Vision*, pp. 69–84, 2016.
- [16] A. van den Oord, Y. Li, and O. Vinyals, "Representation learning with contrastive predictive coding," arXiv preprint arXiv:1807.03748, 2018.
- [17] M. Sabokrou, M. Khalooei, and E. Adeli, "Self-supervised representation learning via neighborhood-relational encoding," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8010– 8019, 2019.
- [18] M. Caron, I. Misra, J. Mairal, P. Goyal, P. Bojanowski, and A. Joulin, "Unsupervised learning of visual features by contrasting cluster assignments," *Advances in Neural Information Processing Systems*, vol. 33, pp. 9912–9924, 2020.
- [19] K. Hassani and A. H. Khasahmadi, "Contrastive multi-view representation learning on graphs," in *International Conference on Machine Learning*, pp. 4116–4126, 2020.
- [20] Y. Tian, D. Krishnan, and P. Isola, "Contrastive multiview coding," in European Conference on Computer Vision, pp. 776–794, 2020.
- [21] R. Couronné, P. Vernhet, and S. Durrleman, "Longitudinal selfsupervision to disentangle inter-patient variability from disease progression," in *International Conference on Medical Image Computing and Computer-Assisted Intervention, Lecture Notes in Computer Science*, vol. 12902, pp. 231–241, 2021.
- [22] I. Higgins *et al.*, "β-VAE: Learning basic visual concepts with a constrained variational framework," in *International Conference on Learning Representations*, 2016.
- [23] R. T. Chen, X. Li, R. Grosse, and D. Duvenaud, "Isolating sources of disentanglement in VAEs," in *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pp. 2615–2625, 2018.
- [24] H. Kim and A. Mnih, "Disentangling by factorising," in *International Conference on Machine Learning*, pp. 2649–2658, 2018.
- [25] W.-N. Hsu, Y. Zhang, and J. Glass, "Unsupervised learning of disentangled and interpretable representations from sequential data," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp. 1876–1887, 2017.
- [26] W. Grathwohl and A. Wilson, "Disentangling space and time in video with hierarchical variational auto-encoders," *arXiv preprint arXiv:1612.04440*, 2016.
- [27] Y. Li and S. Mandt, "Disentangled sequential autoencoder," arXiv preprint arXiv:1803.02991, 2018.
- [28] J. Ouyang et al., "Self-supervised longitudinal neighbourhood embedding," in International Conference on Medical Image Computing and

Computer-Assisted Intervention, Lecture Notes in Computer Science, vol. 12902, pp. 80–89, 2021.

- [29] M. Toepper, "Dissociating normal aging from Alzheimer's disease: A view from cognitive neuroscience," *Journal of Alzheimer's Disease*, vol. 57, no. 2, pp. 331–352, 2017.
- [30] R. Sivera, H. Delingette, M. Lorenzi, X. Pennec, and N. Ayache, "A model of brain morphological changes related to aging and Alzheimer's disease from cross-sectional assessments," *NeuroImage*, vol. 198, pp. 255–270, 2019.
- [31] D. K. Johnson, M. Storandt, J. C. Morris, and J. E. Galvin, "Longitudinal study of the transition from healthy aging to Alzheimer disease," *Archives of Neurology*, vol. 66, no. 10, pp. 1254–1259, 2009.
- [32] S. G. Mueller et al., "Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI)," Alzheimer's & Dementia, vol. 1, no. 1, pp. 55–66, 2005.
- [33] R. C. Petersen *et al.*, "Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization," *Neurology*, vol. 74, no. 3, pp. 201– 209, 2010.
- [34] T. Ohnishi, H. Matsuda, T. Tabira, T. Asada, and M. Uno, "Changes in brain morphology in Alzheimer disease and normal aging: is Alzheimer disease an exaggerated aging process?," *American Journal of Neuroradiology*, vol. 22, no. 9, pp. 1680–1685, 2001.
- [35] M. Lorenzi, X. Pennec, G. B. Frisoni, and N. Ayache, "Disentangling normal aging from Alzheimer's disease in structural magnetic resonance images," *Neurobiology of Aging*, vol. 36, pp. S42–S52, 2015.
- [36] C. R. Jack Jr et al., "The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods," *Journal of Magnetic Resonance Imaging*, vol. 27, no. 4, pp. 685–691, 2008.
- [37] S. G. Mueller *et al.*, "The Alzheimer's Disease Neuroimaging Initiative," *Neuroimaging Clinics*, vol. 15, no. 4, pp. 869–877, 2005.
- [38] H. Li, M. Habes, D. A. Wolk, and Y. Fan, "A deep learning model for early prediction of Alzheimer's disease dementia based on hippocampal magnetic resonance imaging data," *Alzheimer's & Dementia*, vol. 15, no. 8, pp. 1059–1070, 2019.
- [39] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," arXiv preprint arXiv:1312.6114, 2013.
- [40] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *Proceedings of the 26th Annual International Conference* on Machine Learning, pp. 41–48, 2009.
- [41] A. Pentina, V. Sharmanska, and C. H. Lampert, "Curriculum learning of multiple tasks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5492–5500, 2015.
- [42] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.
- [43] J. K. Kueper, M. Speechley, and M. Montero-Odasso, "The Alzheimer's disease assessment scale–cognitive subscale (ADAS-Cog): modifications and responsiveness in pre-dementia populations. a narrative review," *Journal of Alzheimer's Disease*, vol. 63, no. 2, pp. 423–444, 2018.
- [44] D. Tosun et al., "Amyloid status imputed from a multimodal classifier including structural MRI distinguishes progressors from nonprogressors in a mild Alzheimer's disease clinical trial cohort," Alzheimer's & Dementia, vol. 12, no. 9, pp. 977–986, 2016.
- [45] S. M. Landau *et al.*, "Measurement of longitudinal β-amyloid change with 18F-florbetapir PET and standardized uptake value ratios," *Journal* of Nuclear Medicine, vol. 56, no. 4, pp. 567–574, 2015.
- [46] M. Brendel *et al.*, "Improved longitudinal 18F-AV45 amyloid PET by white matter reference and VOI-based partial volume effect correction," *NeuroImage*, vol. 108, pp. 450–459, 2015.
- [47] K. Chen *et al.*, "Improved power for characterizing longitudinal amyloidβ PET changes and evaluating amyloid-modifying treatments with a cerebral white matter reference region," *Journal of Nuclear Medicine*, vol. 56, no. 4, pp. 560–566, 2015.
- [48] "Alzheimer's Disease Neuroimaging Initiative." http://adni. loni.usc.edu/.
- [49] "Florbetapir processing methods." https://adni.bitbucket. io/reference/docs/UCBERKELEYAV45/ADNI_AV45_ Methods_JagustLab_06.25.15.pdf.
- [50] A. D. Joshi *et al.*, "Performance characteristics of amyloid PET with 18F-florbetapir in patients with Alzheimer's disease and cognitively normal subjects," *Journal of Nuclear Medicine*, vol. 53, no. 3, pp. 378– 384, 2012.
- [51] S. Haykin, "A comprehensive foundation," *Neural Networks*, vol. 2, no. 2004, p. 41, 2004.
- [52] A. L. Maas *et al.*, "Rectifier nonlinearities improve neural network acoustic models," in *International Conference on Machine Learning*, vol. 30, pp. 3–8, 2013.

- [53] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann, "The balanced accuracy and its posterior distribution," in 20th International Conference on Pattern Recognition, pp. 3121–3124, 2010.
- [54] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [55] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International Conference on Machine Learning*, pp. 1597–1607, 2020.
- [56] K. Sohn, H. Lee, and X. Yan, "Learning structured output representation using deep conditional generative models," *Advances in Neural Information Processing Systems*, vol. 28, pp. 3483–3491, 2015.
- [57] H. F. Rhodius-Meester *et al.*, "MRI visual ratings of brain atrophy and white matter hyperintensities across the spectrum of cognitive decline are differently affected by age and diagnosis," *Frontiers in Aging Neuroscience*, vol. 9, pp. 117–128, 2017.
- [58] M. Pihlajamaki, A. M. Jauhiainen, and H. Soininen, "Structural and functional MRI in mild cognitive impairment," *Current Alzheimer Research*, vol. 6, no. 2, pp. 179–185, 2009.
- [59] K. Oh, Y.-C. Chung, K. Kim, W.-S. Kim, and I.-S. Oh, "Classification and visualization of Alzheimer's disease using volumetric convolutional neural network and transfer learning," *Scientific Reports*, vol. 9, 2019.
- [60] M. Liu, J. Zhang, E. Adeli, and D. Shen, "Landmark-based deep multiinstance learning for brain disease diagnosis," *Medical Image Analysis*, vol. 43, pp. 157–168, 2018.
- [61] C. Zu et al., "Label-aligned multi-task feature learning for multimodal classification of Alzheimer's disease and mild cognitive impairment," *Brain Imaging and Behavior*, vol. 10, no. 4, pp. 1148–1159, 2016.
- [62] H.-I. Suk, S.-W. Lee, and D. Shen, "Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis," *NeuroImage*, vol. 101, pp. 569–582, 2014.
- [63] W. Lin et al., "Convolutional neural networks-based MRI image analysis for the Alzheimer's disease prediction from mild cognitive impairment," *Frontiers in Neuroscience*, vol. 12, pp. 777–789, 2018.
- [64] Y. Huang, J. Xu, Y. Zhou, T. Tong, and X. Zhuang, "Diagnosis of Alzheimer's disease via multi-modality 3D convolutional neural network," *Frontiers in Neuroscience*, vol. 13, pp. 509–520, 2019.
- [65] T. Zhou, M. Liu, K.-H. Thung, and D. Shen, "Latent representation learning for Alzheimer's disease diagnosis with incomplete multimodality neuroimaging and genetic data," *IEEE Transactions on Medical Imaging*, vol. 38, no. 10, pp. 2411–2422, 2019.
- [66] T. Zhou et al., "Deep multi-modal latent representation learning for automated dementia diagnosis," in *International Conference on Medical Image Computing and Computer-Assisted Intervention, Lecture Notes in Computer Science*, vol. 11767, pp. 629–638, 2019.
- [67] K. R. Gray, R. Wolz, R. A. Heckemann, P. Aljabar, A. Hammers, and D. Rueckert, "Multi-region analysis of longitudinal FDG-PET for the classification of Alzheimer's disease," *NeuroImage*, vol. 60, no. 1, pp. 221–229, 2012.
- [68] C. Platero and M. C. Tobar, "Predicting Alzheimer's conversion in mild cognitive impairment patients using longitudinal neuroimaging and clinical markers," *Brain Imaging and Behavior*, pp. 1–11, 2020.
- [69] T. Niccoli and L. Partridge, "Ageing as a risk factor for disease," *Current Biology*, vol. 22, no. 17, pp. R741–R752, 2012.
- [70] R. Matej, A. Tesar, and R. Rusina, "Alzheimer's disease and other neurodegenerative dementias in comorbidity: a clinical and neuropathological overview," *Clinical Biochemistry*, vol. 73, pp. 26–31, 2019.
- [71] A. Duthie, D. Chew, and R. Soiza, "Non-psychiatric comorbidity associated with Alzheimer's disease," *QJM: An International Journal of Medicine*, vol. 104, no. 11, pp. 913–920, 2011.
- [72] J. A. Santiago and J. A. Potashkin, "The impact of disease comorbidities in Alzheimer's disease," *Frontiers in Aging Neuroscience*, vol. 13, pp. 38–50, 2021.
- [73] R. Moretti, P. Torre, R. M. Antonello, G. Cazzato, and A. Bava, "Depression and Alzheimer's disease: symptom or comorbidity?," *American Journal of Alzheimer's Disease & Other Dementias*, vol. 17, no. 6, pp. 338–344, 2002.

0278-0062 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information. Authorized licensed use limited to: University of Southern California. Downloaded on May 10,2022 at 23:00:27 UTC from IEEE Xplore. Restrictions apply.