

Article

Self-Supervised Contrastive Learning to Predict the Progression of Alzheimer's Disease with 3D Amyloid-PET

Min Gu Kwak ¹, Yi Su ², Kewei Chen ², David Weidman ², Teresa Wu ³, Fleming Lure ⁴,
Jing Li ^{1,*} and for the Alzheimer's Disease Neuroimaging Initiative [†]

¹ H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA; mkwak35@gatech.edu

² Banner Alzheimer's Institute, Phoenix, AZ 85006, USA; yi.su@bannerhealth.com (Y.S.); kewei.chen@bannerhealth.com (K.C.); david.weidman@bannerhealth.com (D.W.)

³ School of Computing and Augmented Intelligence, Arizona State University, Tempe, AZ 85281, USA; teresa.wu@asu.edu

⁴ MS Technologies Corporation, Rockville, MD 20850, USA; fleming.lure@mstechnologies.com

* Correspondence: jli3175@gatech.edu

[†] Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.



Citation: Kwak, M.G.; Su, Y.; Chen, K.; Weidman, D.; Wu, T.; Lure, F.; Li, J.; for the Alzheimer's Disease Neuroimaging Initiative.

Self-Supervised Contrastive Learning to Predict the Progression of Alzheimer's Disease with 3D Amyloid-PET. *Bioengineering* **2023**, *10*, 1141. <https://doi.org/10.3390/bioengineering10101141>

Academic Editors: Alan Wang, Sibisiso Mdletshe, Brady Williamson and Guizhi Xu

Received: 29 August 2023

Revised: 25 September 2023

Accepted: 27 September 2023

Published: 28 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Early diagnosis of Alzheimer's disease (AD) is an important task that facilitates the development of treatment and prevention strategies, and may potentially improve patient outcomes. Neuroimaging has shown great promise, including the amyloid-PET, which measures the accumulation of amyloid plaques in the brain—a hallmark of AD. It is desirable to train end-to-end deep learning models to predict the progression of AD for individuals at early stages based on 3D amyloid-PET. However, commonly used models are trained in a fully supervised learning manner, and they are inevitably biased toward the given label information. To this end, we propose a self-supervised contrastive learning method to accurately predict the conversion to AD for individuals with mild cognitive impairment (MCI) with 3D amyloid-PET. The proposed method, SMoCo, uses both labeled and unlabeled data to capture general semantic representations underlying the images. As the downstream task is given as classification of converters vs. non-converters, unlike the general self-supervised learning problem that aims to generate task-agnostic representations, SMoCo additionally utilizes the label information in the pre-training. To demonstrate the performance of our method, we conducted experiments on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. The results confirmed that the proposed method is capable of providing appropriate data representations, resulting in accurate classification. SMoCo showed the best classification performance over the existing methods, with AUROC = 85.17%, accuracy = 81.09%, sensitivity = 77.39%, and specificity = 82.17%. While SSL has demonstrated great success in other application domains of computer vision, this study provided the initial investigation of using a proposed self-supervised contrastive learning model, SMoCo, to effectively predict MCI conversion to AD based on 3D amyloid-PET.

Keywords: Alzheimer's disease; mild cognitive impairment; amyloid-PET; self-supervised learning; representation learning

1. Introduction

Alzheimer's disease (AD) is a neurodegenerative disease and the most common form of dementia. AD symptoms initially include a loss of short-term memory ability, and as the symptoms become worse, cognitive decline occurs. It is estimated that there are 6.7 million

individuals aged 65 and older affected by AD in the United States alone in 2023 [1]. Despite several decades of unsuccessful drug development, recent times have signaled a glimmer of hope with the full FDA approval of a novel drug, Leqembi [2]. Moreover, another promising medication, donanemab, is under testing, and showing encouraging early results [3]. Notably, these groundbreaking pharmaceutical developments herald a new era in the fight against AD. Yet, their potential to slow disease progression is contingent upon early administration. There is a strong consensus that the most effective treatment regime should target the early stages of the disease before irreversible brain damage has occurred [4]. Thus, the early identification of an individual's condition is important [5].

Mild cognitive impairment (MCI) is a prodromal phase of AD when individuals show noticeable signs of memory and cognitive decline, but the symptoms are not severe enough to disrupt their daily activities. MCI is a high-risk stage that 10~15% of individuals progress to AD each year. It is crucial to identify which MCI individuals will convert to AD (also known as converters), which could provide an opportunity for early intervention to try to slow down the progression. This has been formulated as a classification problem (i.e., classifying MCI individuals into converters vs. non-converters) in the AD literature.

Neuroimaging is an important tool for AD-related assessments, and has demonstrated great potential for predicting MCI conversion to AD. Volumetric magnetic resonance imaging (MRI) and positron emission tomography (PET) are two important neuroimaging modalities [6–9]. MRI can provide information about the structural alteration of the brain [10]. Training end-to-end deep learning models based on 3D neuroimages has focused more on MRI in past research [7,11]. On the other hand, PET can provide information about functional and pathological changes in the brain. A commonly used PET imaging modality is FDG-PET which measures cerebral glucose metabolism. Amyloid-PET is a promising neuroimaging modality for AD diagnosis, as it measures the accumulation of amyloid plaques in the brain—a hallmark of AD. It is of great interest to use amyloid-PET for converter vs. non-converter classification of MCI patients.

There are two limitations of the existing work we want to tackle in this paper: First, the existing studies using amyloid-PET for MCI conversion classification focused on pre-defined features [12] (e.g., regional amyloid measurements). Building a deep learning model that takes the 3D amyloid-PET images as input without feature engineering will greatly complement the existing studies. Second, most existing approaches are supervised learning models which are trained using labeled data only. Labeled samples can be quite limited, especially for training deep learning models with many parameters. Leveraging other available data sources, such as unlabeled data, has great potential to improve model training. This is especially important for training with amyloid-PET, because this imaging modality is not routinely collected for patients, and thus has a much smaller sample size than MRI.

Self-supervised learning (SSL) is a new machine learning paradigm in which a model is trained to learn general representations of input data (e.g., semantic representations of images) with no label information needed. SSL has gained much popularity recently because of its superior capability of learning representations that are broadly transferable to various downstream tasks by fine-tuning, such as image segmentation, object detection, and classification. Using SSL in a pre-training step, the model trained to perform the downstream task can be less biased to the limited labeled data, thus having better generalizability. SSL has resulted in remarkable improvements in various domain applications, including but not limited to natural images [13], histopathology images [14], autonomous driving [15], and medical images [16,17]. Recent studies have especially focused on medical images. The hybrid architecture of UNet and vision transformer, UNETR, was introduced to learn the sequence representations of 3D input for medical image segmentation. It achieved considerable performance gains for multi-organ, brain tumor, and spleen segmentation tasks [18]. Furthermore, UNETR was improved by adopting the Swin Transformer architecture for efficient training. It also introduced several tailored proxy tasks for proper self-supervised learning in the medical domain [19]. However, only a few studies using

SSL in radiology related to AD have been conducted [20]. Furthermore, no study has been performed to predict MCI conversion to AD using 3D amyloid-PET, which motivated our work in this paper.

This study proposes a self-supervised contrastive learning framework, Semi Momentum Contrast (SMoCo), to predict MCI conversion to AD. To the best of our knowledge, it is the first study to leverage the SSL approach for predicting MCI conversion based on 3D amyloid-PET images. We hypothesize that using SSL to obtain general representations from a large amount of unlabeled data can help the model achieve better performance. Acquiring fully labeled datasets in AD research is challenging. The diagnostic process, which requires clinicians evaluating a complex array of information, is both resource-intensive and time-consuming. We address this challenge by employing SSL. Our model is built upon Momentum Contrast (MoCo), a representative existing SSL model that learns representations to minimize a contrastive loss in instance discrimination. To improve MoCo when the downstream task is classification, which is the focus of this paper, we propose SMoCo, which aims to learn more suitable representations for the downstream classification. SMoCo not only leverages unlabeled data, but also exploits label information in the pre-training step. To demonstrate the effectiveness of SMoCo, we conducted experiments on the ADNI dataset and compared it with alternative methods.

2. Materials and Methods

2.1. Data

This retrospective study was conducted on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. ADNI is one of the largest datasets for AD studies to date, with the primary goal being to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. ADNI (<http://adni.loni.ucla.edu>, accessed on 31 October 2022) was launched in 2003 by the NIH, FDA, private pharmaceutical companies, and nonprofit organizations, as a \$60,000,000, 5-year public-private partnership. The primary goal of ADNI has been to test whether MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California-San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the US and Canada. For up-to-date information, please see <http://www.adni-info.org/>, accessed on 31 October 2022.

We downloaded 1064 3D amyloid-PET images from 612 individuals. If an individual, with multiple images taken at different times, converts from MCI to AD within 36 months of an image being captured, that image is labeled as converter. Otherwise, it is labeled as non-converter. There were 158 converters and 463 non-converters, and additionally 443 unlabeled MCI images. The unlabeled MCI refers to images that the AD conversion cannot be determined because there are no records at or after 36 months. The demographic and clinical information of the data are demonstrated in Table 1 and Figure 1.

We also downloaded the T1-weighted MR image corresponding to each PET image. The T1-weighted MR images were spatially normalized using the Computational Anatomy Toolbox 12 [21] with Statistical Parametric Mapping [22] and a standard brain atlas from the Montreal Neurological Institute. Then, each PET image was co-registered with the corresponding MRI. The spatially normalized PET images have a size of $121 \times 145 \times 121$ and a voxel size of 1.5 mm in depth, height, and width. We applied zero padding and resized the images to a size of $72 \times 72 \times 72$ using nearest-neighbor interpolation. The labeled images were split into 80%, 10%, and 10% for training, validation, and testing. All unlabeled images were included in training. The ratio between the labeled and unlabeled data in training is

about 1:0.82. After training and validation, the model was applied to the test set to compute performance metrics. The train/validation/test split was repeated 30 times with different random seeds so that the average and standard deviation of test performance metrics could be reported.

Table 1. Demographic and clinical characteristics of the dataset. The ‘Gender’ column represents the proportion of females, while other values denote the mean with standard deviation in parentheses.

	Gender	Age	Education Years	Mini-Mental State Examination
Converter	42.41%	75.34 (7.51)	15.82 (2.81)	26.97 (2.01)
Non-Converter	42.12%	72.45 (7.59)	16.32 (2.73)	28.33 (1.62)
Unlabeled	39.73%	75.79 (8.16)	16.06 (2.63)	27.82 (2.12)

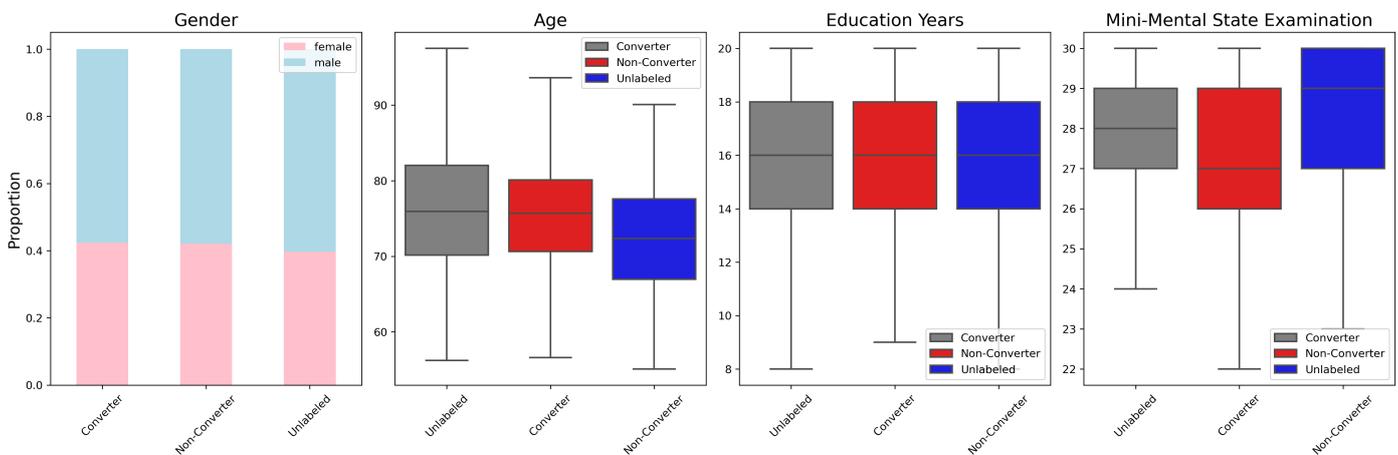


Figure 1. Distributions of demographic and clinical variables in ADNI dataset.

2.2. Proposed SMOCo

We propose a self-supervised contrastive learning method to predict MCI conversion to AD based on 3D amyloid-PET. It is based on the MoCo, a popular SSL method that has recently set a milestone with its great computational efficiency [23,24]. MoCo aims to capture the semantic representations of images in the pre-training step where no label information is needed. Then, the pre-trained network can be fine-tuned to perform various downstream tasks. To further improve MoCo for classification downstream tasks, we introduce SMOCo. SMOCo refines pre-training representations for classification by leveraging a contrastive loss function that incorporates label information. In the following, we will briefly review MoCo and then introduce SMOCo.

Let $\mathcal{D} = \mathcal{D}_L \cup \mathcal{D}_U$ be a training dataset, where \mathcal{D}_L and \mathcal{D}_U denote the labeled and unlabeled amyloid-PET images, respectively. In the pre-training step, MoCo trains a network by discarding the label information and learning semantic representations of the images through instance discrimination. Formally, given an image $x_i \in \mathcal{D}$, a stochastic data augmentation $t(\cdot)$ is applied to the same image twice to generate two different views, $x_i^a = t(x_i)$ and $x_i^+ = t(x_i)$, called the anchor and a positive instance, respectively. By putting x_i^a and x_i^+ through a query network $f_\theta(\cdot)$ and a key network $f_\phi(\cdot)$ with shared structure, we can obtain their respective representation vectors, $z_i^a = f_\theta(x_i^a)$, $z_i^+ = f_\phi(x_i^+)$, respectively. We should train the network to make z_i^a and z_i^+ similar, i.e., to “pull” the positive instance toward the anchor because both of them are generated from the same image. In addition, we can draw K images other than x_i from the training set and apply augmentation $t(\cdot)$ to these images to obtain $\{x_{ik}^-\}_{k=1}^K$, which are called negative instances. By putting each x_{ik}^- through $f_\phi(\cdot)$, we can obtain its representation vector, $z_{ik}^- = f_\phi(x_{ik}^-)$. We should train the network to make z_i^a and z_{ik}^- dissimilar, i.e., to “push” each negative instance away from the

anchor because they are different images. To realize the ideas of the “pull” and “push”, MoCo uses the following loss function:

$$\mathcal{L}_i^{MoCo} = -\log \frac{\exp(z_i^a \cdot z_i^+ / \tau)}{\exp(z_i^a \cdot z_i^+ / \tau) + \sum_{k=1}^K \exp(z_i^a \cdot z_{ik}^- / \tau)}. \tag{1}$$

τ is a temperature hyperparameter for scaling. Under this loss function, MoCo trains the query network $f_\theta(\cdot)$ and the key network $f_\phi(\cdot)$ with the same structure, but updating parameters ϕ by an exponential moving average of θ . Also, MoCo uses a memory queue to store past representations of negative instances to save computational costs.

Note that MoCo does not use label information, even though the training set \mathcal{D} includes a subset of labeled samples \mathcal{D}_L . Leveraging the label information has the potential to learn semantic representations that are more appropriate for the downstream classification task. To achieve this, we propose SMOCo, which “pulls” additional instances toward the anchor. These are instances in the memory queue with the same label as the anchor. Formally, for each labeled image $x_i \in \mathcal{D}_L$, recall that the anchor is obtained by applying augmentation to the image, i.e., $x_i^a = t(x_i)$. $\{z_{im}^-\}_{m=1}^{M_i} \subset \{z_{ik}^-\}_{k=1}^K$ denote a subset of M_i instances in the memory queue of length K which have the same label as the anchor. To “pull” these instances toward the anchor, we propose the following loss:

$$\mathcal{L}_i^{Label} = -\frac{1}{|M(i; t)|} \log \sum_{m \in M_i} \frac{\exp(z_i^a \cdot z_{im}^+ / \tau)}{\exp(z_i^a \cdot z_{im}^+ / \tau) + \sum_{k=1}^K \exp(z_i^a \cdot z_{ik}^- / \tau)}. \tag{2}$$

The final SMOCo loss function is defined as a combined loss of (1) and (2):

$$\mathcal{L}^{SMoCo} = \sum_{i \in \mathcal{D}} \mathcal{L}_i^{MoCo} + \alpha \sum_{i \in \mathcal{D}_L} \mathcal{L}_i^{Label}, \tag{3}$$

where α is a balancing hyperparameter. It is worth mentioning that, although it may be possible to pull more instances toward the anchor, e.g., by relying on some pseudo-labels of unlabeled images, we chose a more conservative approach in SMOCo based only on labeled images. This is to avoid introducing labeling noise to the learning of semantic representations. Figure 2 provides a graphical overview of SMOCo.

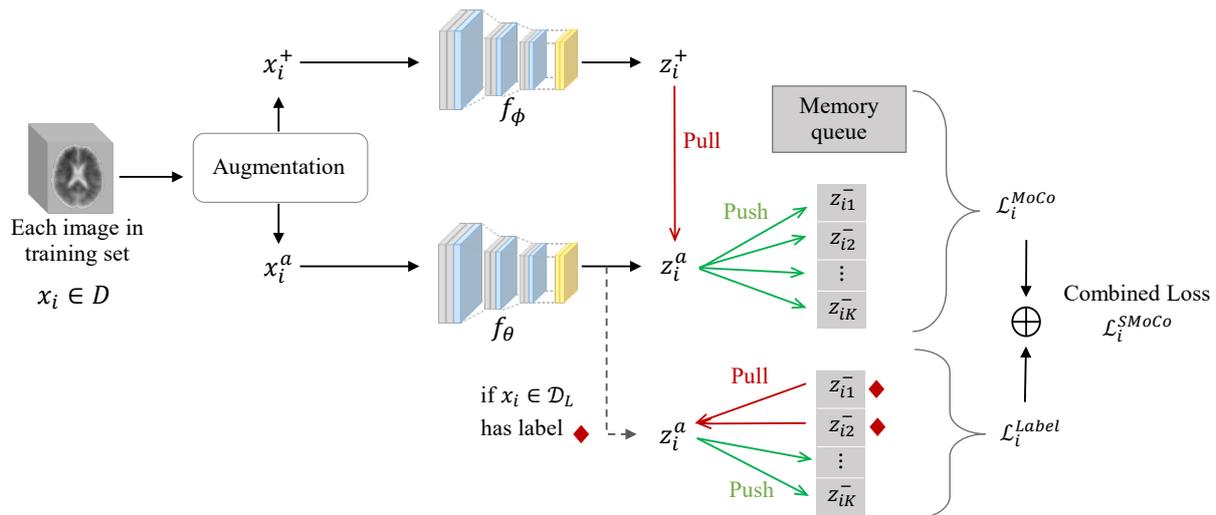


Figure 2. Graphical overview of SMOCo. For a given image x_i , two augmentations are applied to generate a positive instance x_i^+ and an anchor x_i^a . Both instances are fed into 3D ResNet-50 encoders f_ϕ and f_θ to obtain representations z_i^+ and z_i^a , respectively. \mathcal{L}_i^{MoCo} aims to pull z_i^+ toward z_i^a because they are created from the same image, while pushing other instances in the memory queue away from z_i^a . \mathcal{L}_i^{Label} leverages label information from the memory queue, ensuring the representations from the same class are pulled closer to z_i^a . \mathcal{L}_i^{MoCo} and \mathcal{L}_i^{Label} are combined as the final loss in SMOCo, \mathcal{L}_i^{SMoCo} .

2.3. SMoCo Implementation Details and Fine-Tuning

Inspired by recent works [23–25], a 3D ResNet-50 encoder [26] with the fully connected layers replaced by a two-layer multilayer perceptron was chosen for the key and query networks. For the encoder, we replaced the first $7 \times 7 \times 7$ convolution layer by a $3 \times 3 \times 3$ convolution layer with a stride of one and zero padding of one. The architecture of the 3D ResNet-50 encoder is depicted in Figure 3. The output dimension of the multilayer perceptron was set to 128.

We set $\tau = 0.2, K = 1024$, and the exponential moving average coefficient as 0.95. Pre-training was performed for 100 epochs with a batch size of 16. The network was optimized using the AdamW [27] optimizer with a momentum of 0.9 and a learning rate of 0.0001. The learning rate is gradually dropped to zero by following a half-cosine schedule. For the proposed SMoCo loss function, values in 0.25, 0.5, 1, 2, 3, 5 were used to investigate the effect of the balancing hyperparameter α .

After the pre-training is finished, the next step is fine-tuning. Specifically, the multi-layer perceptron of the query network was substituted with a single-layer softmax classifier. Then, the network was trained using the pre-trained weights as initial values to minimize the cross-entropy loss for classification based on \mathcal{D}_L . During inference, this fine-tuned network was used to predict the conversion status of each new patient based on their 3D amyloid-PET. All models were implemented with Pytorch [28] and scikit-learn [29].

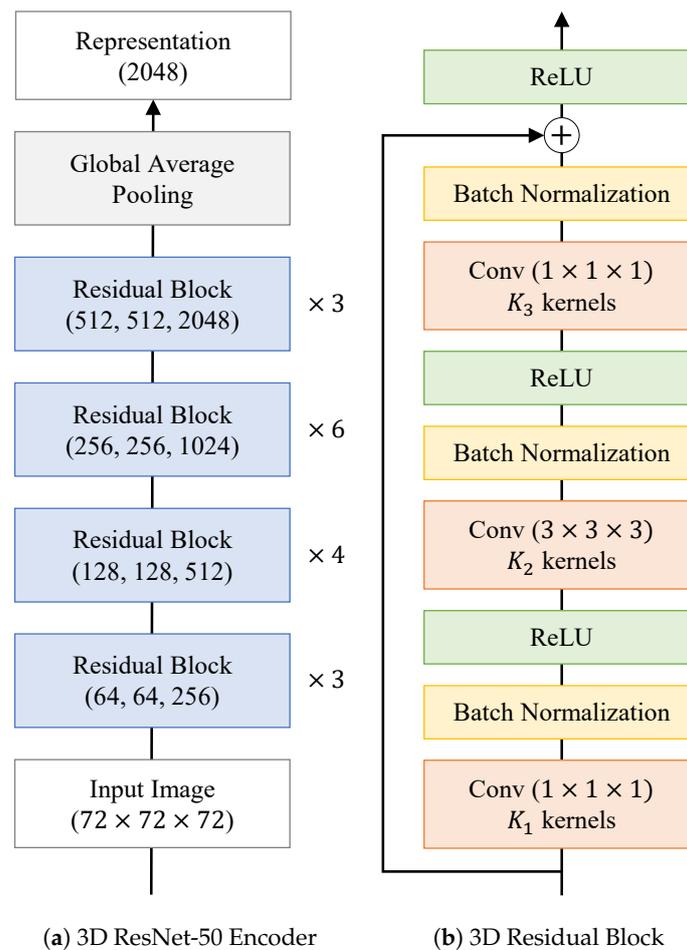


Figure 3. (a) Structure of ResNet-50 encoder used for SMoCo (the same encoder is used for f_ϕ and f_θ). The numbers in a bracket denote K_1, K_2 , and K_3 of a 3D residual block, respectively. (b) Structure of 3D residual block in the encoder.

3. Results

3.1. Representation Quality Evaluation for Pre-Training Step

In SSL, high-quality representations learned in the pre-training step are important for the downstream task. Here, we compared the representation quality of SMOCo and MoCo. Specifically, we trained SMOCo to minimize the loss in Equation (3) with $\alpha = 1$ and obtained the representation vector for each training sample. The same was performed for MoCo. To visualize the distribution of the training samples, we used Uniform Manifold Approximation and Projection (UMAP) to reduce the dimensionality of the representation vector to two. Figure 4 compares the UMAP representations of MoCo and SMOCo. As shown in Figure 4b, converter and non-converter samples overlapped more when MoCo was used. During learning of the semantic structure of the images, converters and non-converters were separated to some extent because they have different amyloid-PET characteristics. However, the classes were not separated enough, because MoCo is designed to produce general representations, not for the specific classification task. Compared with Figure 4a, we can confirm that SMOCo more clearly separated converters and non-converters.

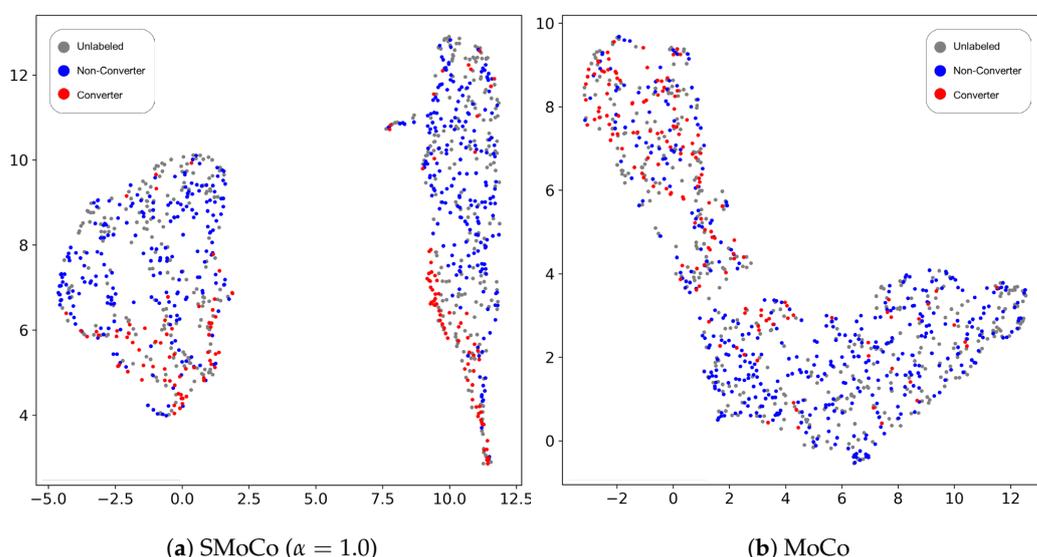


Figure 4. UMAP visualization of the representations of training images. (a) SMOCo; (b) MoCo. Grey, blue, and red points refer to the unlabeled images, converters, and non-converters, respectively.

Furthermore, we compared SMOCo and MoCo using a more quantitative approach than visualization. The idea was that a better representation should entail a better classification of labeled samples based on their representation vectors. To this end, we obtained the representation vectors of samples in the validation set by applying the trained SMOCo and MoCo. Different values of the hyperparameter α were tried for SMOCo. Then, a k -nearest neighbor (k -NN) algorithm with $k = 5$ was used to classify each validation sample, and the Area under the Receiver Operating Characteristics (AUROC) was reported to appropriately evaluate the models with class imbalance. k -NN was adopted because it has been a common choice to evaluate representation quality of SSL [30]. As shown in Table 2, it can be observed that SMOCo enhances the representation quality over MoCo regardless of the value of α . Namely, the proposed loss function helps the model to provide more appropriate representations for classification tasks because, it is designed to perform the given purpose well. The best AUROC is obtained when $\alpha = 1$, which is 4.70% higher than MoCo. Furthermore, it can be observed that the model performance gradually decreases when α is greater than one. Our interpretation is that the model focused too much on aggregating the instances of the same class so that the instance discrimination task was not properly performed. Overall, a good balance between the two losses combined in SMOCo, i.e., \mathcal{L}^{MoCo} and \mathcal{L}^{Label} , is important to achieve the best representation for dif-

ferentiating converters vs. non-converters. Based on the results, we fixed $\alpha = 1$ in the remaining experiments.

Table 2. Representation quality comparison in the pre-training step on validation data. The average values with standard deviations are reported. The best result is **boldfaced**.

Model	MoCo			SMoCo			
α	0	0.25	0.5	1	2	3	5
AUROC	76.37 (3.60)	79.24 (3.32)	80.35 (2.94)	81.07 (3.02)	80.63 (3.71)	79.41 (3.40)	78.96 (3.55)

Moreover, we showed the AUROC on validation data along training epochs for both MoCo and SMoCo in Figure 5. Both models demonstrate stable convergence. SMoCo consistently outperforms MoCo with higher AUROC across the training epochs. The SMoCo curve increases earlier than MoCo, indicating that the integration of label information during pre-training can accelerate the model's ability to capture representations vital for downstream classification.

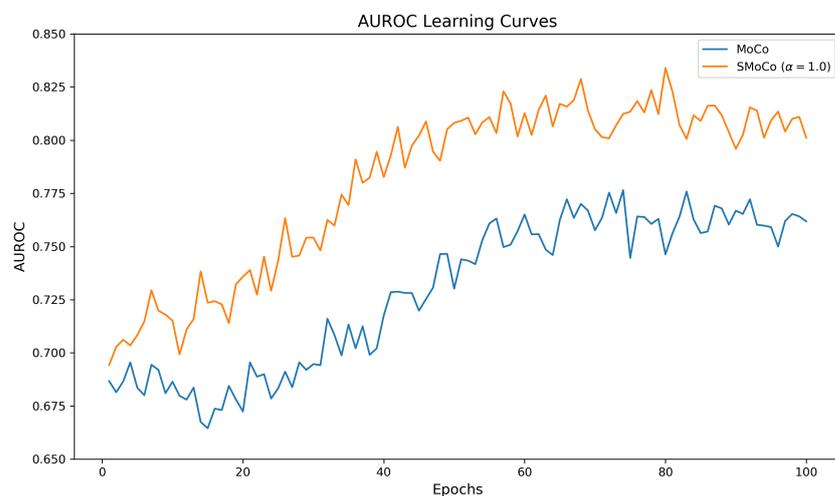


Figure 5. Comparing AUROC of SMoCo and MoCo across training epochs. SMoCo shows faster and efficient training, as well as higher performance than MoCo.

3.2. Classification Performance and Comparison

Finally, we conducted extensive experiments comparing the classification accuracy of our method with a variety of existing methods, including supervised classification, three popular semi-supervised learning methods (pseudo-labeling [31], virtual adversarial training [32], and stochastic weight averaging [33]), and MoCo.

Supervised classification refers to the conventional model that is trained using only labeled data \mathcal{D}_L . It was trained for 100 epochs with an AdamW optimizer using an initial learning rate of 0.0001. The learning rate was also decreased to zero using a half-cosine schedule. The batch size was set to 16. Semi-supervised learning models were trained on \mathcal{D}_L and \mathcal{D}_U . Unlike SSL, which involves a pre-training and a fine-tuning step, these models were trained at once by incorporating both the cross-entropy loss for \mathcal{D}_L and an additional loss for \mathcal{D}_U proposed by the corresponding method. The same training hyperparameters as supervised classification were used.

To evaluate the classification performance of MoCo and SMoCo, we fine-tuned the networks for 10 epochs. Other training hyperparameters were kept the same as supervised classification. These methods are referred to as “MoCo and Fine-Tuning” and “SMoCo and Fine-Tuning” in Table 3. In addition, recognizing that fine-tuning the entire network requires considerable time, another commonly used approach is to use an SSL model as a

feature extractor and train a simple classifier based on the representations [34]. We used this approach and trained a random forest classifier [35] with the generated representation vectors from the pre-trained SMOCo. This method is referred to as “SMoCo and Random Forest” in Table 3. Note that random forest cannot be used directly on 3D image data and a prior feature extraction step is needed. Therefore, the “SMoCo and Random Forest” approach is intended to demonstrate the effectiveness of SMOCo as a feature extractor.

Furthermore, noting that our amyloid-PET data has a class imbalance, we applied a resampling technique in every mini-batch to address the problem. We also used an adaptive cutoff strategy to select the threshold for classification probability to assign each sample into a binary class (converters vs. non-converters), which is a recommended strategy under class imbalance [36]. We reported evaluation metrics such as AUROC, accuracy, sensitivity, and specificity.

Table 3 presents the classification results. Overall, classification performance is good in the order of SMOCo, MoCo, semi-supervised methods, and supervised classification. It can be noticed that simply using MoCo enhanced the model performances, especially in terms of AUROC and sensitivity. It improved the prediction of the minority class (converters). We can confirm that the general data representations learned from SSL led to classification performance improvements.

Table 3. Classification performance on test data. The average values with standard deviations are reported. The best result is **boldfaced**.

Category	Model	AUROC	Accuracy	Sensitivity	Specificity
Supervised	Supervised Classification	81.53 (3.81)	77.68 (4.01)	73.20 (4.22)	78.89 (3.65)
Semi-Supervised	Pseudo-Labeling	81.89 (3.93)	77.97 (3.53)	73.22 (3.68)	79.18 (3.97)
	Virtual Adversarial Training	82.03 (3.36)	78.13 (3.99)	73.43 (2.98)	78.03 (3.50)
	Stochastic Weight Averaging	82.27 (3.88)	78.19 (3.45)	73.65 (3.39)	78.08 (4.10)
Self-Supervised	MoCo and Fine-Tuning	83.01 (3.59)	78.37 (3.13)	74.23 (2.89)	78.39 (3.77)
	SMoCo and Random Forest	84.86 (3.31)	79.10 (3.09)	74.96 (3.58)	80.03 (3.12)
	SMoCo and Fine-Tuning	85.17 (2.87)	81.09 (3.38)	77.39 (2.97)	82.17 (3.26)

Moreover, the proposed SMOCo further improved the performance of MoCo. Both training a random forest classifier and applying fine-tuning showed better performance than other models. Fine-tuned SMOCo achieved the best performance with considerable gains of 2.16%p of AUROC, 2.72%p of accuracy, 3.27%p of sensitivity, and 3.78%p of specificity than MoCo. This, in turn, proves our original conjecture that “pulling” additional instances with the same label as the labeled images in the training set, as performed by SMOCo, help learn more suitable representations for the downstream classification and bring a substantial performance gap.

4. Discussion

Compared to the other application areas of computer vision, the use of SSL in AD studies is quite limited, with only a handful of recent papers focusing on 3D MRI. To classify AD and health controls, an SSL method utilizing a data augmentation technique that mixes medically relevant regions was proposed [37]; a multimodal SSL model was used to combine structural and functional MRI [20]. To predict MCI conversion, a benchmarking study was performed, which revealed that some SSL methods have advantages over supervised pre-training, multitask learning, and multiclass learning [38]. Also, a two-stage model was proposed, which combined transfer learning and self-supervised contrastive learning [11]. However, no study using SSL on 3D amyloid-PET has been found.

On the other hand, there are existing studies using amyloid-PET, by itself or combined with other imaging modalities, for MCI conversion prediction. However, these studies are based on pre-extracted features. One study used the fractal dimension and Shannon entropy as extracted features from amyloid-PET and trained a support vector machine for

classification [12]. Another study built a multimodal sparse representation-based classifier based on pre-defined features from various regions of interest obtained from amyloid-PET and MRI [39]. A transfer learning method was proposed to exploit features extracted from regions of interest of amyloid-PET, FDG-PET, and MRI, which can account for missing modalities [40]. In comparison to these studies, our study achieved better and similar performance, but provided an end-to-end method based directly on 3D amyloid-PET without feature engineering.

This study has several limitations. First, SMoCo tends to pull negative instances of the same class as the anchor from the memory queue. This might make the model over-rely on class information, potentially ignoring subtle within-class variations. Recognizing that a class can have diverse patterns, as seen when diagnosing patients with varied symptoms under the same label, a more nuanced approach is required. Like Prototypical Contrastive Learning [41], we can address this by clustering instances and identifying a representative ‘prototype’ for each cluster. This can allow model to account for both the nuances of individual instances and the broader class patterns. Second, this study is based on amyloid-PET data only. Integrating other data modalities such as demographics, clinical records, and MRI has the potential to improve model performance. In a proof-of-concept experiment, we added age, gender, education years, and mini-mental state examination to the fine-tuning stage of SMoCo based on amyloid-PET, as demonstrated in Table 1 and Figure 1. We found improvement over using amyloid-PET alone (AUROC = 86.12%, accuracy = 82.19%, sensitivity = 78.52%, and specificity = 83.71%). We expect further improvement by including MRI, which can be explored in future research. Third, it is well-known that training deep learning models requires large amounts of data, while the sample size of our study is still limited. To expand the training capacity, we could leverage pre-trained 3D networks based on large medical image datasets such as Swin UNETR [42] and Med3D [43]. Last but not least, this study is based on ADNI data. It is important to further validate the proposed method using other datasets than ADNI. To this end, we acquired another public dataset commonly used for AD studies, the Australian Imaging Biomarkers and Lifestyle Study of Ageing (AIBL) [44]. The AIBL dataset contains a limited number of amyloid-PET images with 28 converters and 14 non-converters. To apply our method to the AIBL dataset, we further fine-tuned the previously obtained ADNI-based model using AIBL data. We used 5-fold cross validation, i.e., including 4 folds of AIBL data to further fine-tune the ADNI-based model, testing on the remaining fold, and iterating this process through all folds to compute performance. This resulted in an AUROC of 82.50% on the AIBL dataset, which is comparable to the testing performance using the ADNI dataset (AUROC = 85.17%). This result demonstrated the generalizability of our method, while we acknowledge that the AIBL dataset has a limited sample size. Further validation using larger datasets is needed, and will be explored in future research.

5. Conclusions

In summary, our study is among the first ones that leverage SSL to predict MCI conversion to AD based on 3D amyloid-PET Images. Amyloid-PET images have favorable characteristics for early AD diagnosis. We used 3D images to avoid using feature engineering that requires domain knowledge and related tools. The main advantage of SSL is to enable the leveraging of a large amount of unlabeled images to learn general representations, which helps improve the downstream classification task. In AD research, collecting a sufficient amount of diagnosis labels is costly and time-consuming. Therefore, utilizing unlabeled data can be an important benefit. To further strengthen the advantage of SSL, we proposed SMoCo to learn more suitable representations for the downstream classification task of converters and non-converters. Our experimental results showed that SMoCo outperformed a variety of existing SSL, semi-supervised learning, and supervised learning models.

Author Contributions: Conceptualization, M.G.K. and J.L.; methodology, M.G.K.; validation, M.G.K.; formal analysis, M.G.K.; investigation, M.G.K.; resources, J.L.; data curation, Y.S., K.C. and D.W.; writing—original draft preparation, M.G.K. and J.L.; writing—review and editing, D.W., T.W. and F.L.; visualization, M.G.K.; supervision, J.L.; funding acquisition, J.L.; software, M.G.K.; project administration, J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by NIH grant 2R42AG053149-02A1 and NSF grant DMS-2053170. This research was also supported by NIH grants R01AG069453 and P30AG072980, the State of Arizona, and Banner Alzheimer’s Foundation. Data collection and sharing for this project was funded by the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer’s Association; Alzheimer’s Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org, accessed on 31 October 2022). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer’s Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data used in the preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in the analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf. The ADNI dataset analyzed in this research work is available at <https://adni.loni.usc.edu/> (accessed on 21 August 2021).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AD	Alzheimer’s disease
ADNI	Alzheimer’s Disease Neuroimaging Initiative
AIBL	Australian Imaging Biomarkers and Lifestyle Study of Ageing
AUROC	Area under the receiver operating characteristic
<i>k</i> -NN	<i>k</i> -nearest neighbor
MCI	Mild cognitive impairment
MoCo	Momentum contrast
MRI	Magnetic resonance imaging
PET	Positron emission tomography
SSL	Self-supervised learning
UMAP	Uniform manifold approximation and projection

References

1. Alzheimer’s Association. 2023 Alzheimer’s disease facts and figures. *Alzheimer’s Dement. J. Alzheimer’s Assoc.* **2023**, *19*, 1598–1695. [[CrossRef](#)]
2. Canady, V.A. FDA approves new treatment for Alzheimer’s disease. *Ment. Health Wkly.* **2023**, *33*, 6–7.

3. Sims, J.R.; Zimmer, J.A.; Evans, C.D.; Lu, M.; Ardayfio, P.; Sparks, J.; Wessels, A.M.; Shcherbinin, S.; Wang, H.; Nery, E.S.M.; et al. Donanemab in early symptomatic Alzheimer disease: The TRAILBLAZER-ALZ 2 randomized clinical trial. *JAMA* **2023**, *330*, 512–527.
4. Cummings, J.; Lee, G.; Ritter, A.; Sabbagh, M.; Zhong, K. Alzheimer’s disease drug development pipeline: 2020. *Alzheimer’s Dement. Transl. Res. Clin. Interv.* **2020**, *6*, e12050.
5. Battineni, G.; Chintalapudi, N.; Hossain, M.A.; Losco, G.; Ruocco, C.; Sagaro, G.G.; Traini, E.; Nittari, G.; Amenta, F. Artificial intelligence models in the diagnosis of adult-onset dementia disorders: A review. *Bioengineering* **2022**, *9*, 370.
6. Illakiya, T.; Ramamurthy, K.; Siddharth, M.; Mishra, R.; Udainiya, A. AHANet: Adaptive Hybrid Attention Network for Alzheimer’s Disease Classification Using Brain Magnetic Resonance Imaging. *Bioengineering* **2023**, *10*, 714.
7. Katabathula, S.; Wang, Q.; Xu, R. Predict Alzheimer’s disease using hippocampus MRI data: A lightweight 3D deep convolutional network model with visual and global shape representations. *Alzheimer’s Res. Ther.* **2021**, *13*, 1–9.
8. Lu, D.; Popuri, K.; Ding, G.W.; Balachandar, R.; Beg, M.F. Multimodal and multiscale deep neural networks for the early diagnosis of Alzheimer’s disease using structural MR and FDG-PET images. *Sci. Rep.* **2018**, *8*, 5697.
9. Zhang, D.; Wang, Y.; Zhou, L.; Yuan, H.; Shen, D.; Initiative, A.D.N.; Alzheimer’s Disease Neuroimaging Initiative. Multimodal classification of Alzheimer’s disease and mild cognitive impairment. *Neuroimage* **2011**, *55*, 856–867.
10. Ritchie, K.; Lovestone, S. The dementias. *Lancet* **2002**, *360*, 1759–1766.
11. Lu, P.; Hu, L.; Zhang, N.; Liang, H.; Tian, T.; Lu, L. A two-stage model for predicting mild cognitive impairment to Alzheimer’s disease conversion. *Front. Aging Neurosci.* **2022**, *14*, 826622. [[CrossRef](#)]
12. Yan, Y.; Somer, E.; Grau, V. Classification of amyloid PET images using novel features for early diagnosis of Alzheimer’s disease and mild cognitive impairment conversion. *Nucl. Med. Commun.* **2019**, *40*, 242–248. [[CrossRef](#)]
13. Caron, M.; Touvron, H.; Misra, I.; Jégou, H.; Mairal, J.; Bojanowski, P.; Joulin, A. Emerging properties in self-supervised vision transformers. In Proceedings of the IEEE/CVF International Conference on Computer Vision 2021, Montreal, BC, Canada, 11–17 October 2021; pp. 9650–9660.
14. Ciga, O.; Xu, T.; Martel, A.L. Self supervised contrastive learning for digital histopathology. *Mach. Learn. Appl.* **2022**, *7*, 100198. [[CrossRef](#)]
15. Luo, C.; Yang, X.; Yuille, A. Self-supervised pillar motion learning for autonomous driving. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2021, Nashville, TN, USA, 19–25 June 2021; pp. 3183–3192.
16. Chowdhury, A.; Rosenthal, J.; Waring, J.; Umeton, R. Applying self-supervised learning to medicine: Review of the state of the art and medical implementations. *Informatics* **2021**, *8*, 59. [[CrossRef](#)]
17. Xing, X.; Liang, G.; Wang, C.; Jacobs, N.; Lin, A.L. Self-Supervised Learning Application on COVID-19 Chest X-ray Image Classification Using Masked AutoEncoder. *Bioengineering* **2023**, *10*, 901. [[CrossRef](#)]
18. Hatamizadeh, A.; Tang, Y.; Nath, V.; Yang, D.; Myronenko, A.; Landman, B.; Roth, H.R.; Xu, D. Unetr: Transformers for 3d medical image segmentation. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision 2022, Waikoloa, HI, USA, 3–8 January 2022; pp. 574–584.
19. Tang, Y.; Yang, D.; Li, W.; Roth, H.R.; Landman, B.; Xu, D.; Nath, V.; Hatamizadeh, A. Self-supervised pre-training of swin transformers for 3d medical image analysis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2022, New Orleans, LA, USA, 18–24 June 2022; pp. 20730–20740.
20. Fedorov, A.; Wu, L.; Sylvain, T.; Luck, M.; DeRamus, T.P.; Bleklov, D.; Plis, S.M.; Calhoun, V.D. On self-supervised multimodal representation learning: An application to Alzheimer’s disease. In Proceedings of the 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), Nice, France, 13–16 April 2021; pp. 1548–1552.
21. Gaser, C.; Dahnke, R.; Thompson, P.M.; Kurth, F.; Luders, E.; Initiative, A.D.N. CAT—A computational anatomy toolbox for the analysis of structural MRI data. *bioRxiv* **2022**. [[CrossRef](#)]
22. Ashburner, J.; Barnes, G.; Chen, C.C.; Daunizeau, J.; Flandin, G.; Friston, K.; Kiebel, S.; Kilner, J.; Litvak, V.; Moran, R.; et al. *SPM12 Manual*; Wellcome Trust Centre for Neuroimaging: London, UK, 2014; Volume 2464.
23. Chen, X.; Fan, H.; Girshick, R.; He, K. Improved baselines with momentum contrastive learning. *arXiv* **2020**, arXiv:2003.04297.
24. He, K.; Fan, H.; Wu, Y.; Xie, S.; Girshick, R. Momentum contrast for unsupervised visual representation learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2020, Seattle, WA, USA, 14–19 June 2020; pp. 9729–9738.
25. Caron, M.; Misra, I.; Mairal, J.; Goyal, P.; Bojanowski, P.; Joulin, A. Unsupervised learning of visual features by contrasting cluster assignments. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 9912–9924.
26. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2016, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
27. Loshchilov, I.; Hutter, F. Decoupled weight decay regularization. *arXiv* **2017**, arXiv:1711.05101.
28. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. Pytorch: An imperative style, high-performance deep learning library. *Adv. Neural Inf. Process. Syst.* **2019**, *32*.
29. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
30. Wu, Z.; Xiong, Y.; Yu, S.X.; Lin, D. Unsupervised feature learning via non-parametric instance discrimination. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2018, Salt Lake City, UT, USA, 18–22 June 2018; pp. 3733–3742.

31. Lee, D.H. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In Proceedings of the Workshop on Challenges in Representation Learning, ICML, Atlanta, GA, USA, 16–21 June 2013; Volume 3, p. 896.
32. Miyato, T.; Maeda, S.i.; Koyama, M.; Ishii, S. Virtual adversarial training: A regularization method for supervised and semi-supervised learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *41*, 1979–1993. [[CrossRef](#)]
33. Athiwaratkun, B.; Finzi, M.; Izmailov, P.; Wilson, A.G. There are many consistent explanations of unlabeled data: Why you should average. *arXiv* **2018**, arXiv:1806.05594.
34. Purushwalkam, S.; Gupta, A. Demystifying contrastive self-supervised learning: Invariances, augmentations and dataset biases. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 3407–3418.
35. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
36. Hajian-Tilaki, K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Casp. J. Intern. Med.* **2013**, *4*, 627.
37. Seyfioğlu, M.S.; Liu, Z.; Kamath, P.; Gangolli, S.; Wang, S.; Grabowski, T.; Shapiro, L. Brain-Aware Replacements for Supervised Contrastive Learning in Detection of Alzheimer’s Disease. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention 2022, Singapore, 18–22 September 2022; pp. 461–470.
38. Jiang, H.; Miao, C. Pre-Training 3D Convolutional Neural Networks for Prodromal Alzheimer’s Disease Classification. In Proceedings of the 2022 International Joint Conference on Neural Networks (IJCNN), Padua, Italy, 18–23 July 2022; pp. 1–8.
39. Xu, L.; Wu, X.; Li, R.; Chen, K.; Long, Z.; Zhang, J.; Guo, X.; Yao, L.; Alzheimer’s Disease Neuroimaging Initiative. Prediction of progressive mild cognitive impairment by multi-modal neuroimaging biomarkers. *J. Alzheimer’s Dis.* **2016**, *51*, 1045–1056. [[CrossRef](#)] [[PubMed](#)]
40. Liu, X.; Chen, K.; Weidman, D.; Wu, T.; Lure, F.; Li, J.; Initiative, A.D.N. A novel transfer learning model for predictive analytics using incomplete multimodality data. *IIEE Trans.* **2021**, *53*, 1010–1022. [[CrossRef](#)] [[PubMed](#)]
41. Li, J.; Zhou, P.; Xiong, C.; Hoi, S.C. Prototypical contrastive learning of unsupervised representations. *arXiv* **2020**, arXiv:2005.04966.
42. Hatamizadeh, A.; Nath, V.; Tang, Y.; Yang, D.; Roth, H.R.; Xu, D. Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In Proceedings of the International MICCAI Brainlesion Workshop 2021, Virtual Event, 27 September 2021; pp. 272–284.
43. Chen, S.; Ma, K.; Zheng, Y. Med3d: Transfer learning for 3d medical image analysis. *arXiv* **2019**, arXiv:1904.00625.
44. Ellis, K.A.; Bush, A.I.; Darby, D.; De Fazio, D.; Foster, J.; Hudson, P.; Lautenschlager, N.T.; Lenzo, N.; Martins, R.N.; Maruff, P.; et al. The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: Methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer’s disease. *Int. Psychogeriatr.* **2009**, *21*, 672–687. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.