Cross-Domain Self-Supervised Deep Learning for Robust Alzheimer’s Disease Progression Modeling

Saba Dadsetan\textsuperscript{1} \hspace{1cm} Mohsen Hejrati\textsuperscript{2} \hspace{1cm} Shandong Wu\textsuperscript{1}\hspace{1cm} Somaye Hashemifar\textsuperscript{2}

\textsuperscript{1}University of Pittsburgh \hspace{1cm} \textsuperscript{2}Genentech, Inc.

sabadad@cs.pitt.edu, hashems4@gene.com

Abstract

Developing successful artificial intelligence systems in practice depends both on robust deep learning models as well as large high quality data. Acquiring and labeling data can become prohibitively expensive and time-consuming in many real-world applications such as clinical disease models. Self-supervised learning has demonstrated great potential in increasing model accuracy and robustness in small data regimes. In addition, many clinical imaging and disease modeling applications rely heavily on regression of continuous quantities. However, the applicability of self-supervised learning for these medical-imaging regression tasks has not been extensively studied. In this study, we develop a cross-domain self-supervised learning approach for disease prognostic modeling as a regression problem using 3D images as input. We demonstrate that self-supervised pre-training can improve the prediction of Alzheimer’s Disease progression from brain MRI. We also show that pre-training on extended (but not labeled) brain MRI data outperforms pre-training on natural images. We further observe that the highest performance is achieved when both natural images and extended brain-MRI data are used for pre-training.

1. Introduction

Developing reliable and robust artificial intelligence systems requires advanced and efficient deep learning techniques, as well as the creation and annotation of large volumes of data for training. However, the construction of a labeled dataset is often time-consuming and expensive, such as in the medical imaging domain, due to the complexity of annotation tasks and the high expertise required for the manual interpretation. To alleviate the lack of annotations in medical imaging, transfer learning from natural images has been successful and has demonstrated great potential in many downstream applications in computer vision, where the labeling process is quite expensive and time-consuming \textsuperscript{[4,15,20,27,44,46,56,63,65]}. Medical research and healthcare are especially well-poised to benefit from self-supervised learning approaches, given the prevalence of unprecedented amounts of medical images generated by hospital and non-hospital settings. Despite demand, the use of self-supervised approaches in the medical image domain has received limited attention. Only few studies have investigated the impact of self-supervised learning in the medical image analysis domain for limited applications including classification \textsuperscript{[2,25,32,34,53,68]} and segmentation \textsuperscript{[5,9,50,54]}.

In this study, we focus on developing a self-supervised deep learning model for predicting the progression of Alzheimer’s disease (AD) by using high dimensional magnetic resonance imaging (MRI). AD is a slowly progressing disease caused by the degeneration of brain cells, with patients showing clinical symptoms years after the onset of the disease. Therefore, accurate prognosis and treatment of AD in its early stage is critical to prevent non-reversible and fatal brain damage.

Many of the existing methods have focused on the task of classifying patients into coarse categories: Cognitive Normal (CN), Mild Cognitive Impairment (MCI), or Alzheimer’s Disease (AD), or to predict conversion from one class to another (e.g. MCI to AD) \textsuperscript{[45,49,59]}. However, applications in clinical trials require a more fine-grained measurement scale, because clinical trial populations are typically narrowly defined (eg. only MCI). Another approach is to instead predict the outcome of cognitive and functional tests such as Clinical Dementia Rating Scale Sum of Boxes (CDR-SB), which are measured by continuous numerical values. Prognostic prediction, modeled as a regression task rather than a classification, helps provide more granular estimates of disease progression. Recently, a few deep learning based approaches, including the recurrent
neural network (RNN) and convolutional neural networks (CNN) have been proposed for predicting disease progression of AD patients based on MRIs. [43] adapted MiniRNN to integrate longitudinal clinical information and cross-sectional tabular imaging features for regressing endpoints. [17] utilized an ensemble model based on stacked CNN and a bidirectional long short-term memory (BiLSTM) to predict the endpoints on the fusion of time series clinical features and derived imaging features. Recently, several methods have started to employ CNN based models to extract features from raw medical imaging. [55] applied CNN with a multi-task interaction layers composed of feature decoupling modules and feature interaction module to predict the disease progression.

Despite demand, little progress has been made because of the difficult design requirements, lack of large-scale, homogeneous datasets that contain early stage AD patients, and noisy endpoints that are potentially hard to predict. Much of the prior work has focused on using image-derived features (using FreeSurfer, ANTS) in order to overcome the complexity and high variability in raw MRI images, and small dataset sizes. Most of the current prognosis models are trained a single dataset (i.e., cohort), which cannot be easily generalized to other cohorts. Moreover, all of them only utilize limited number of annotated images, which may suffer from problems arising from domain shift and heterogeneity. Moreover, a large number of parameters in CNN are involved in training on a small dataset, which could lead to over-fitting [16].

We establish a cross-domain self-supervised transfer learning approach for medical image analysis that leverages self-supervised learning on both unlabeled large-scale natural images and a domain-specific medical imaging from 11 different internal and external studies to learn transferable and generalizable representations for medical images. Such representations can be further fine-tuned and deployed for downstream tasks such as disease progression prediction using limited labeled data from the clinical setting. Our transfer learning strategy benefits from both large-scale self-supervised representations and domain-specific self-supervised representations in a unified deep learning framework. We develop a detailed and extensive framework to evaluate the performance of our model fine-tuned from different supervised and self-supervised models pre-trained on either natural images or medical images, or both. Our extensive experimental results reveal that (1) self-supervised pre-training on natural images followed by intermediate self-supervised learning on unlabeled domains-specific medical data outperforms all other transfer learning techniques, which is very promising because replacing supervised approaches with self-supervised learning can significantly reduce the need for data annotation (2) self-supervised models pre-trained on medical images outperform same models pre-trained on natural images indicating that self-supervised learning on medical images yield discriminative feature representations for regression task.

2. Related works

The recent development and success of self-supervised learning techniques, including contrastive learning [10–12, 22, 23, 39, 61], mutual information reduction [57], clustering [8, 31], and redundancy-reduction methods [6, 64] in computer vision indicate their effectiveness in improving the performance of AI systems. These methods train models on different pretext tasks to enable the network to learn high-quality representations without label information. SimCLR [12] maximizes agreement between representations of different augmentations of the same image by using a contrastive loss in the latent space. Barlow Twins [64] measure the cross-correlation matrix between the embedding of two identical networks and its goal is to make this cross-correlation close to the identity matrix. SwAV [8] simultaneously clusters the images while enforcing consistency between cluster assignments produced for differently augmented views of the same image, instead of comparing features directly as in contrastive learning.

Subsequently, self-supervised learning has been employed for medical imaging applications including classification and segmentation to learn visual representations of medical images by incorporating unlabeled medical images. While some approaches have designed domain-specific pretext tasks [5, 54, 68, 69], others have adjusted well-known self-supervised learning methods to medical data [25, 30, 53, 66]. Very recently [2] has applied SimCLR on a combination of unlabeled ImageNet dataset and task specific medical images for medical image classification; their experiments and improved performance suggest that pre-training on ImageNet is complementary to pre-training on unlabeled medical images.

Although aforementioned approaches demonstrate improvement of the performance on challenging medical datasets, all of them are limited to classification and segmentation tasks and their benefits and potential effects for the prognosis prediction tasks, as regression tasks, have not been studied. Formulating progress prediction as a regression rather than a traditional classification problem leads to a more fine-grained measurement scale which is crucial for real-world applications. Therefore, the development of self-supervised networks is in great demand for efficient data-utilization in medical imaging for disease prognosis. To the best of our knowledge, this is the first study of developing a self-supervised deep convolution neural network on medical data images from various cross-domain datasets to predict a granular understanding of disease progression.
3. Methodology

3.1. Task and Data

In Alzheimer’s Disease clinical trials, the most important cognitive test that is performed to assess current patient function and the likelihood of AD progression is Clinical Dementia Rating Scale Sum of Boxes (CDR-SB). Our goal is to predict the progression of Alzheimer’s Disease 12 months after the first patient visit. Specifically, our model uses 2D slice stacks of a magnetic resonance imaging (MRI) volume (which are collected at the first visit) as an input and predicts the CDR-SB value, which is computed at month 12 (i.e., after one year). All individuals included in our analysis are around 86 from seven external datasets, including ADNI [47], BIOFINDER [36], FACEHBI [41], AIBL [18], HABs [13], BIOCARD [40], and WRAP [29], and four internal studies including A, B, C, and D (dataset names are anonymized in order to allow for blind review). We only include the axial view of MR images for this paper analysis. MR volumes are standardized using standard preprocessing steps (in supplementary materials) and are cropped or padded to \((224 \times 224)\) to make them compatible with the ResNet-50 input size. Furthermore, using the predefined SynthSeg model [7], we generate segmented images from raw images in order to evaluate the information of different brain parts. The segmentation results only contain the brain region and masked out other details such as skull, etc. The results of segmented images are provided in the supplementary materials.

We provide two datasets of medical images referred to as center-slice and 5-slice datasets. In the 5-slice dataset, for each subject, we prepare a 5-slice stack from its corresponding 3D MRI volume by taking the center slices of the brain sub-volumes and four adjacent slices at intervals equal to 5 (two to the right, two to the left). The center-slice imaging dataset only contains the center slice, while the 5-slice dataset includes the 5-slice stack. The number of unlabeled images in the center-slice dataset is 43,740, and the number of unlabeled images in the 5-slice dataset equals 218,700. Both of these datasets are used to train self-supervised models. During the self-supervised learning process, 92% and 8% of the data are assigned for the training and validation, respectively. Following the common practice in self-supervised learning, based on the minimum self-supervised validation loss, we select the best model and transfer its backbone weights to the supervised model. For the purpose of creating labeled datasets, we use MR images of participant’s first visits. The participants are selected from five datasets, including ADNI, A, B, C, and D. All datasets were filtered to include patients who are positive for amyloid pathology, have a MMSE larger than 20 and are diagnosed with early stages of AD i.e. prodromal or mild at baseline. Amyloid positivity is detected by either cerebrospinal fluid (CSF) or amyloid positron emission tomography (PET). For supervised training, we only consider the 2D center slice. The number of labeled data in this dataset is 935 for the train and validation sets, as well as 360 for the independent test set.

3.2. Self-supervised learning platform for progress prediction problem

Following are the experiments that comprise our approach. 1) We assign our model with self-supervised pre-trained representations on unlabeled natural images in ImageNet. 2) We initialize the weights of model randomly then do self-supervised pre-training on medical images (Figure 1(a)). 3) We initialize weights with supervised pre-trained weights on labeled ImageNet and then self-supervised pre-training on medical images (Figure 1(b)). 4) We initialize the weights with self-supervised pre-trained weights on unlabeled ImageNet and then self-supervised pre-training on all layers on unlabeled medical images (Figure 1(c)). A number of data augmentations have been performed for all the self-supervised learning models discussed in this paper, including random gray scaling, random cropping with resizing, random Gaussian blurring, random rotation, and random flipping. We use a batch size of 50 and epoch of 1000 and implement an early-stop mechanism based on the validation data in order to prevent over-fitting. After we have pre-trained the weights of self-supervised frameworks, we use their CNN backbone (in our experiments, this is ResNet-50) to fine-tune the regression task (Figure 1(d)) and then calculate the results on an independent set of data (Figure 1(e)). For supervised learning models, we perform data augmentation including random vertical and horizontal rotation, as well as random flipping. For supervised learning, a batch size of 25 and an epoch of 500 are used.

Self-supervised pre-training on natural images

There are a wide variety of self-supervised learning (SSL) models that are pre-trained on large-scale natural image datasets, such as ImageNet, that are commonly used for transfer learning. Such SSL pre-training techniques surpass supervised ImageNet models in several computer vision tasks. To exploit these benefits, in this experiment, we initialize the backbone encoder with weights from SSL models trained on ImageNet. While several types of self-supervised learning methods exist, in this work, we focus on three self-supervised learning methods, one contrastive approach including SimCLR, one redundancy reduction approach including BarLow Twins (BLT), and one clustering-based contrastive learning approach including SwAV which have yielded impressive results on natural image benchmarks. There are several other self-supervised learning strategies, but their performance on ImageNet yields comparable results.
Self-supervised pre-training on medical images

Although representations learned from natural images perform well on medical imaging tasks [42], they may not be optimal for the medical imaging domain due to the large distribution shift from natural images. Generally, medical images are monochromatic and have similar anatomical structures. This discrepancy could be minimized by further pre-training on medical data, either supervised or unsupervised. Due to the time and cost associated with the annotation of medical data, self-supervised learning can be considered a more practical and realistic alternative. Similar to self-supervised pre-training on natural images, we employ SimCLR, Barlow Twins, and SwAV to learn distinctive representations of unlabeled medical images effectively. These methods have all demonstrated good performance in the classification and segmentation of medical images.

Fine Tuning

During fine-tuning, we train the model using the weights of the pre-trained network as initialization for the downstream supervised task. For progress prediction, the standard ResNet-50 backbone [24] is followed by a linear layer head, where the backbone is initialized either randomly or with the pre-trained models. Training ResNet-50 involves freezing backbone parameters and only training the head. The mean square error (MSE) is used as a criterion for calculating model loss.

4. Results

We propose the first benchmarking analysis to evaluate the efficacy of different pre-training models for disease progression prediction as a regression problem. (1) We examine the transferability of both supervised and self-supervised ImageNet models pre-trained on natural images for the purpose of assessing their performance on medical tasks; (2) we investigate the efficacy of pre-training models on domain specific medical images for transfer learning performance on medical tasks; and (3) we evaluate cross-domain self-supervised pre-training on both natural and medical datasets to explore the efficacy of the proposed platform for bridging the domain gap between natural and medical images. In order to evaluate the performance of our model, we used the coefficient of determination, which is indicated as $R^2$.

$$R^2 = 1 - \frac{\sum (e_i^2)}{\sum (y_i - \bar{y})^2} \tag{1}$$

where $\sum (e_i^2)$ shows that sum of squared residuals and $\sum (y_i - \bar{y})^2$ shows total sum squared. $R^2$ is commonly used in clinical studies to assess how well a model explains and predicts future outcomes [19]. However, most of the current literature report pearson correlation or minimum squared error for evaluating their models which alone is not an indicative of ”goodness of fit”.

4.1. Self-supervised models on natural images provide more generalizable representations

Experimental setup. This experiment evaluates the transferability of standard supervised ImageNet models with three popular SSL methods using officially released models including SimCLR, Barlow Twins, and SwAV. All SSL models are pre-trained on the ImageNet dataset and employ a ResNet-50 backbone. To establish a baseline, we train the model using random initialization (i.e., without pre-training).

Results. Table 1a displays the results, from which we draw the following indications: (1) transfer learning from the supervised ImageNet model does not improve over random initialization. We attribute this performance to the significant domain shift between the pre-training and regression target task. In particular, supervised ImageNet models tend to capture domain-specific semantic features, which can be inefficient if the pre-training and target data distributions differ. This finding is consistent with recent studies on other medical tasks that demonstrate that transfer learning from supervised ImageNet pre-training does not always correlate with performance on either classification or segmentation of medical images [14, 28, 58]. (2) Transfer learning from self-supervised ImageNet models provides superior performance compared with both random initialization and transfer learning from the supervised ImageNet model. The best self-supervised model (i.e., Barlow Twins) achieves performance improvement of 7% and 8% over random initialization and the supervised ImageNet model, respectively. As opposed to supervised pre-trained models, self-supervised pre-trained models, encode semantic features that are not biased toward any particular task, which can improve generalizability cross domains.

4.2. Self-supervised pre-training on medical images outperforms ImageNet models

Experimental setup. We investigate the impact of pre-training medical datasets on self-supervised learning. We train three SSL methods, SimCLR, Barlow Twins, and SwAV on 11 cross-domain unlabeled medical imaging datasets (referred to as in-domain dataset). All of the SSL models are initialized randomly and later are fine tuned by using our labeled dataset.

Results. Table 1b shows the performance accuracy measured by the $R^2$ when SSL methods are pre-trained using the center slice dataset. We observe that (1) SimCLR pre-training on in-domain dataset achieves the highest performance and provides 7% and 4% performance boost compared to SwAV and Barlow Twins respectively. This observation may offer the superiority of contrastive learning for...
identifying significant MRI features for predicting progression of Alzheimer’s disease in terms of CDRSB. (2) SimCLR pre-training on in-domain dataset yields higher performance than either supervised or self-supervised pre-training on only the ImageNet dataset (as seen in Table 1a). Intuitively, pre-training on in-domain dataset encodes domain-specific features that reflect unique characteristics of medical images.

Table 1. Comparison of different pre-training schemes on downstream task.

| Pre-Training    | Initialization  | $R^2$ |
|-----------------|-----------------|-------|
| -               | Random          | 0.07  |
| Supervised      | ImageNet        | 0.06  |
| Self-Supervised | SimCLR          | 0.10  |
|                 | SwAV            | 0.08  |
|                 | Barlow Twins    | 0.14  |

(a) Pre-Training on Natural Images.

| Pre-Training    | Initialization  | $R^2$ |
|-----------------|-----------------|-------|
| -               | Random          | 0.07  |
| Supervised      | SimCLR          | 0.17  |
| Self-Supervised | SwAV            | 0.10  |
|                 | Barlow Twins    | 0.13  |

(b) Pre-Training on Medical Images.

The Barlow Twins pre-training on in-domain data, however, does not improve the performance over the Barlow Twins pre-training on ImageNet. This performance suggests that features learned by Barlow Twins pre-trained on ImageNet are generalizable enough to medical images; And therefore pre-training on a limited number of unlabeled images compared with the ImageNet dataset (40k vs. 1.3 M) may offer limited performance gain for Barlow Twins method, which is a redundancy reductions based model.
4.3. Impact of the size of the unlabeled in-domain dataset

Experimental setup. We conduct further experiments to evaluate the advantages of increasing the number of unlabeled images for self-supervised pre-training. For each one of SSL methods, we train two separate models by pre-training on either center-slice or 5-slice datasets.

Results. As observed in Table 2, all three SSL models pre-training on 5-slice imaging dataset results in an improvement over pre-training on center-slice dataset, indicating that the SSL models can benefit from larger pre-training in-domain unlabeled datasets. Specifically, Barlow Twins pre-training on the 5-slice dataset provides 1% performance improvement over pre-training on center-slice dataset which may confirm our previous assumption that Barlow Twins, in contrast to SimCLR, needs more unlabeled in-domain images to constrain the ImageNet-based features for medical tasks.

| Self-Supervised Model | Dataset     | $R^2$ |
|-----------------------|-------------|-------|
| SimCLR                | center-slice| 0.17  |
|                       | 5-slice     | 0.19  |
| SwA/V                 | center-slice| 0.10  |
|                       | 5-slice     | 0.12  |
| Barlow Twins          | center-slice| 0.13  |
|                       | 5-slice     | 0.14  |

Table 2. A comparison of the performance of self-supervised pre-training models in the fine-tuning step with datasets of different sizes.

4.4. Cross-domain self-supervised learning bridges the domain gap between natural and medical images

Experimental setup. We investigate the impact of self-supervised pre-training on both natural images and domain-specific medical images. To do so, we pre-train SimCLR on our 5-slice in-domain dataset with two different initialization schemes including (1) Supervised ImageNet (referred to as Labeled_ImageNet→In-domain), (2) Barlow Twins on the ImageNet dataset (referred to as Unlabeled_ImageNet→In-domain). We select SimCLR and Barlow Twins because as indicated in Tables 1a and 1b, they provide the best results on natural and medical images.

We fine-tune all pre-trained models for the progress prediction task using the labeled 1k images in the training data. Figure 1 shows the cross-domain self-supervised learning employed for the regression problem. In order to show the relation between predicted values and target value we also calculate another metric Pearson correlation coefficient ($r$) 2 metrics for our best performance model and randomly initialized model.

$$r = \frac{\text{cov}(X, Y)}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2(y_i - \bar{y})^2}}$$ (2)

Results. The results are shown in Table 3a. We observe that the best performance is achieved when both unlabeled ImageNet and in-domain datasets are utilized for pre-training. Specifically, Unlabeled_ImageNet→In-domain pre-training surpasses both in-domain and ImageNet pre-trained best performing models. It achieve 14%, 7%, and 2% performance boosts compared to Random initialization, pre-training only on ImageNet, and pre-training only on in-domain dataset, respectively. These results, in line with previous studies [3, 28], imply that pre-training on ImageNet is complementary to pre-training on in-domain datasets, resulting in more powerful representations for medical applications. We also note that Labeled_ImageNet→In-domain pre-training is inferior to Unlabeled_ImageNet→In-domain. These observations restate the efficacy of self-supervised models in delivering more generic representations that can be used for target tasks with limited data, resulting in reduced annotation costs. Supplementary materials also includes the analysis of these cross-domain experiments but for center-slice dataset. Figure 2 illustrate a scatter plot comparing the target CDRSB value with the predicted CDRSB value using a supervised model. As shown in these plots, the red line represents the perfectly fitted model for target values, while the blue line represents the actual fitted model for predicted values. In Figure 2a we use randomly initialized pre-training model and the model shows the results of $r$ of 0.29. Figure 2b illustrates results of the pre-training method Barlow-twin on Natural images, which demonstrates a $r$ value of 0.39. In Figure 2c, we use our best-performing pre-training model, Unlabeled_ImageNet→In-domain, which shows the results of $r$ of 0.46. This higher number compared to randomly initialized model is also illustrated by the decrease incline of blue line towards the red line.

4.5. Performance on test set

Experimental setup. The robustness of our best model, i.e. Barlow Twins→SimCLR on the 5-slice dataset is evaluated and compared with best performing models initialized by either ImageNet or in-domain dataset. Our test set includes 360 unseen patient images.

Results. As presented in Table 3a, the best performance is achieved when both unlabeled ImageNet and in-domain dataset are utilized for pre-training. In particular Unlabeled_ImageNet→In-domain pre-training yields a significant improvement of 10% and 6% over in-domain and ImageNet pre-trained models, respectively. These results confirm our previous observation that supervised pre-
trained models on natural images encode semantic features that are generalizable across domains. Moreover, Labeled\_ImageNet$\rightarrow$In-domain pre-training is inferior to Unlabeled\_ImageNet$\rightarrow$In-domain indicating that supervised ImageNet models encode domain-specific semantic features, which may be inefficient when the pre-training and target data distributions are far apart.

### 4.6. The saliency map shows interpretable features from the cross-domain self-supervised learning model

** Experimental setup.** Attribution methods are a tool for investigating and validating machine learning models. Using the interpretability of the ML models, can significantly help obtaining a bigger picture about risk factors influ-
| Pre-Training Method       | Pre-Training Dataset          | $R^2$ |
|--------------------------|-------------------------------|-------|
| Random                   | -                             | 0.07  |
| Barlow Twins             | ImageNet                      | 0.14  |
| SimCLR                   | In-domain                     | 0.19  |
| Barlow Twins $\rightarrow$ SimCLR | Unlabeled ImageNet $\rightarrow$ In-domain | 0.21  |
| Supervised ImageNet $\rightarrow$ SimCLR | Labeled ImageNet $\rightarrow$ In-domain | 0.18  |

(a) Best performing model in each domain and their combination in cross-domain self-supervised learning setting

| Pre-Training Method       | Pre-Training Dataset          | $R^2$ on Independent test set |
|--------------------------|-------------------------------|-------------------------------|
| Random                   | -                             | 0.04                          |
| Barlow Twins             | ImageNet                      | 0.06                          |
| SimCLR                   | In-domain                     | 0.10                          |
| Barlow Twins $\rightarrow$ SimCLR | Unlabeled ImageNet $\rightarrow$ In-domain | 0.16  |
| Supervised ImageNet $\rightarrow$ SimCLR | Labeled ImageNet $\rightarrow$ In-domain | 0.12  |

(b) Results on independent dataset

Table 3. Results of different pre-training schemes on both (a) validation $R^2$ and (b) independent test $R^2$

ences on short-term prognosis. We used the GradCAM [52] method to extract and evaluate the varying importance of each part of brain MRIs using a gradient of final score. In this method, regions of an image are marked with different colors ranging from red to blue. Generally, the closer the color is to the red color, the greater the contribution from that area to the final result based on the current input data (i.e., MRI slice).

**Results.** Figure 3 illustrates several examples of saliency map. These saliency maps reflect the importance of different regions from the MRIs in pixel level. A general observation we can make through these saliency maps is that our best model i.e. Barlow Twins $\rightarrow$ SimCLR highlights the subcortical areas of the brain. Our observation agrees with previous findings, that the earliest stages of AD show accumulation of abnormal tau in entorhinal cortex and subcortical brain regions [33, 51]. Therefore, it is understandable that our model i.e. Barlow Twins $\rightarrow$ SimCLR gives higher attention to those regions of the brain in a population with patients with prodromal and mild Alzheimer’s disease at baseline. The model initialized randomly highlights random features all around the MRIs and even in background areas, while in model initialized from unsupervised ImageNet, target regions of brain contribute more than other areas (i.e. eye, background) for Alzheimer’s disease prognosis.

**5. Discussion**

We present a cross-domain self-supervised learning framework for predicting the progression of Alzheimer’s disease, which is formulated as a regression problem. This formulation helps to provide more granular estimations of disease progression but can be very noisy and subjective across studies – making it challenging to establish treatment efficacy in small patient populations. The goal of this problem is to predict CDRSB score as an endpoint for defining the stage of Alzheimer’s disease. This framework is enforced to use pre-training on natural images followed by pre-training on domain-specific medical images to alleviate domain shift and improve generalization of pre-trained features. Our extensive experiments demonstrate the effectiveness of our approach to combat the lack of large-scale annotated data for training deep models for progression prediction. Our results show an improvement in prediction accuracy and more robust prediction performance for patients with Alzheimer’s disease. The proposed framework has the potential to identify patients at higher risk of progressing to AD and help develop better therapies at lower cost to society. Because the number of annotated images is relatively small, our current platform only explores the advantages of 2D MRIs for prognostic prediction which decreases parameters and leads to model simplicity. 3D deep neural networks require many more learnable parameters to solve such a complex problem and do not always yield proper outcomes [60]. However, as future works, we will utilize 3D MRIs to compare their performance versus 2D slices. we also will expand our dataset to include more than 5-slice from each image, as well as other views such as coronals and sagittals, allowing us to introduce more in-domain information to our pre-training model.
References

[1] Laith Alzubaidi, Mohammed A Fadhel, Omran Al-Shamma, Jinglan Zhang, J Santamaría, Ye Duan, and Sameer R. Oleivi. Towards a better understanding of transfer learning for medical imaging: a case study. Applied Sciences, 10(13):4523, 2020.

[2] Shekoofeh Azizi, Laura Culp, Jan Freyberg, Basil Mustafa, Sebastien Baur, Simon Kornblith, Ting Chen, Patricia MacWilliams, S Sara Mahdavi, Ellery Wulczyn, et al. Robust and efficient medical imaging with self-supervision. arXiv preprint arXiv:2205.09723, 2022.

[3] Shekoofeh Azizi, Basil Mustafa, Fiona Ryan, Zachary Beaver, Jan Freyberg, Jonathan Deaton, Aaron Loh, Alan Karthikesalingam, Simon Kornblith, Ting Chen, et al. Big self-supervised models advance medical image classification. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 3478–3488, 2021.

[4] Philip Bachman, R Devon Hjelm, and William Buchwalter. Learning representations by maximizing mutual information across views. Advances in neural information processing systems, 32, 2019.

[5] Wenjia Bai, Chen Chen, Giacomo Tarroni, Jinming Duan, Adrien Bardes, Jean Ponce, and Yann LeCun. Variance-invariance-covariance regularization for self-supervised learning. arXiv preprint arXiv:2105.04906, 2021.

[6] Benjamin Billot, Douglas N Greve, Oula Puonti, Axel Thielscher, Koen Van Leemput, Bruce Fischl, Adrian V Dalca, and Juan Eugenio Iglesias. Synbseg: Domain randomisation for segmentation of brain mri scans of any contrast and resolution. arXiv preprint arXiv:2107.09559, 2021.

[7] Matthilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised learning of visual features by contrasting cluster assignments. Advances in Neural Information Processing Systems, 33:9912–9924, 2020.

[8] Krishna Chaitanya, Ertunc Erdil, Neerav Karani, and Ender Konukoglu. Contrastive learning of global and local features for medical image segmentation with limited annotations. Advances in Neural Information Processing Systems, 33:12546–12558, 2020.

[9] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In International conference on machine learning, pages 1597–1607. PMLR, 2020.

[10] Ting Chen, Simon Kornblith, Kevin Swersky, Mohammad Norouzi, and Geoffrey E Hinton. Big self-supervised models are strong semi-supervised learners. Advances in neural information processing systems, 33:22243–22255, 2020.

[11] Xinlei Chen, Haoqi Fan, Ross Girshick, and Kaiming He. Improved baselines with momentum contrastive learning. arXiv preprint arXiv:2003.04297, 2020.

[12] Alexander Dagley, Molly LaPoint, Willem Huijbers, Trey Hedden, Donald G McLaren, Jasmeer P Chatwal, Kathryn V Papp, Rebecca E Amariglio, Deborah Blacker, Dorene M Rentz, et al. Harvard aging brain study: dataset and accessibility. Neuroimage, 144:255–258, 2017.

[13] Jonas Dippel, Steffen Vogler, and Johannes Hönne. Towards fine-grained visual representations by combining contrastive learning with image reconstruction and attention-weighted pooling. arXiv preprint arXiv:2104.04323, 2021.

[14] Carl Doersch, Abhinav Gupta, and Alexei A Efros. Unsupervised visual representation learning by context prediction. In Proceedings of the IEEE international conference on computer vision, pages 1422–1430, 2015.

[15] Amir Ebrahimi, Suhuai Luo, Alzheimer’s Disease Neuroimaging Initiative, et al. Convolutional neural networks for Alzheimer’s disease detection on mri images. Journal of Medical Imaging, 8(2):024503, 2021.

[16] Shaker El-Sappagh, Tamer Abuheimed, SM Riazul Islam, and Kyung Sup Kwak. Multimodal multitask deep learning model for Alzheimer’s disease progression detection based on time series data. Neurocomputing, 412:197–215, 2020.

[17] JR Ellis, Pradeep Jonathan Nathan, Victor L Villemagne, RS Mulligan, Timothy Saudner, K Young, Catherine L Smith, J Welch, Michael Woodward, Keith A Wesnes, et al. Galantamine-induced improvements in cognitive function are not related to alterations in α4/7 nicotinic receptors in early Alzheimer’s disease as measured in vivo by 2-[18F]fluoro-a-85380 PET. Psychopharmacology, 202(1):501–511, 2020.

[18] Nicolai Franzmeier, Nikolaos Koutsouleris, Tammyne Benziger, Alison Goate, Celeste M Karch, Anne M Fagan, Eric McDade, Marco Duering, Martin Dichgans, Johannes Levin, et al. Predicting sporadic Alzheimer’s disease progression via inherited Alzheimer’s disease-informed machine learning. Alzheimer’s & Dementia, 16(3):501–511, 2020.

[19] Spyros Gidaris, Praveer Singh, and Nikos Komodakis. Unsupervised representation learning by predicting image rotations. arXiv preprint arXiv:1803.07728, 2018.

[20] Mara Graziani, Vincent Andrearczyk, and Henning Müller. Visualizing and interpreting feature reuse of pretrained cnns for histopathology. In Irish Machine Vision and Image Processing Conference (IMVIP 2019), Dublin, Ireland, 2019.

[21] Jean-Bastien Grill, Florian Strub, Florent Altch, Corentin Tallec, Pierre Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Guo, Mohammad Ghezelaghli Azar, et al. Bootstrap your own latent—a new approach to self-supervised learning. Advances in neural information processing systems, 33:21271–21284, 2020.

[22] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 9729–9738, 2020.
[24] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016. 4
[25] Xuehai He, Xingyi Yang, Shanghang Zhang, Jinyu Zhao, Yichen Zhang, Eric Xing, and Pengtao Xie. Sample-efficient deep learning for covid-19 diagnosis based on ct scans. medrxiv, 2020.
[26] Michal Heker and Hayit Greenspan. Joint liver lesion segmentation and classification via transfer learning. arXiv preprint arXiv:2004.12352, 2020.
[27] Olivier Henaff. Data-efficient image recognition with contrastive predictive coding. In International conference on machine learning, pages 4182–4192. PMLR, 2020.
[28] Mohammad Reza Hosseinzadeh Taher, Fatemeh Haghighi, Olivier Henaff, Hongwei Li, Fei-Fei Xue, Krishna Chaitanya, Shengda Luo, Jingyu Liu, Gangming Zhao, Yu Fei, Ming Zhang, Yizhou Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Junnan Li, Pan Zhou, Caiming Xiong, and Steven CH Hoi. Knowledge transfer for melanoma screening with deep learning. In 2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017), pages 297–300. IEEE, 2017.
[29] Ishan Misra and Laurens van der Maaten. Self-supervised learning of pretext-invariant representations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 6707–6717, 2020. 2
[30] Hongwei Li, Fei-Fei Xue, Krishna Chaitanya, Shengda Luo, Ivan Ezhev, Benedikt Wiestler, Jianguo Zhang, and Bjoern Menze. Imbalance-aware self-supervised learning for 3d radiomic representations. In International Conference on Medical Image Computing and Computer-Assisted Intervention, pages 36–46. Springer, 2021.
[31] Junnan Li, Pan Zhou, Caiming Xiong, and Steven CH Hoi. Prototypical contrastive learning of unsupervised representations. arXiv preprint arXiv:2005.04966, 2020. 2
[32] Jingyu Liu, Gangming Zhao, Yu Fei, Ming Zhang, Yizhou Wang, and Yizhou Yu. Align, attend and locate: Chest x-ray diagnosis via contrast induced attention network with limited supervision. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 10632–10641, 2019.
[33] Li Liu, Valerie Drouet, Jessica W Wu, Menno P Wit ter, Scott A Small, Catherine Clelland, and Karen Duff. Trans-synaptic spread of tau pathology in vivo. PloS one, 7(2):e31302, 2012. 8
[34] Quande Liu, Lequan Yu, Luyang Luo, Qi Dou, and Pheng Ann Heng. Semi-supervised medical image classification with relation-driven self-ensembling model. IEEE transactions on medical imaging, 39(11):3429–3440, 2020. 1
[35] Yuan Liu, Ayush Jain, Clara Eng, David H Way, Kang Lee, Peggy Bui, Kimberly Kanada, Guillerme de Oliveira Marinho, Jessica Gallegos, Sara Gabriele, et al. A deep learning system for differential diagnosis of skin diseases. Nature medicine, 26(6):900–908, 2020. 1
[36] Niklas Mattsson-Carlsgen, Antoine Leuzy, Sharena Janelidze, Sebastian Palmqvist, Erik Stromrud, Olof Stranberg, Ruben Smith, and Oskar Hansson. The implications of different approaches to define at (n) in alzheimer disease. Neurology, 94(21):e2233–e2244, 2020. 3
[37] Niklas Mattsson-Carlgren, Antoine Leuzy, Sharena Janelidze, Sebastian Palmqvist, Erik Stromrud, Olof Stranberg, Ruben Smith, and Oskar Hansson. The implications of different approaches to define at (n) in alzheimer disease. Neurology, 94(21):e2233–e2244, 2020. 3
[38] Scott Mayer McKinney, Marcin Sieniek, Varun Godbole, Jonathan Godwin, Natasha Antropova, Hutan Ashrafi an, Trevor Back, Mary Chesus, Greg S Corrado, Ara Darzi, et al. International evaluation of an ai system for breast cancer screening. Nature, 577(7788):89–94, 2020. 1
[39] Afonso Menegola, Michel Fornaciali, Ramon Pires, Flávia Vasques Bittencourt, Sandra Avila, and Eduardo Valle. Knowledge transfer for melanoma screening with deep learning. In 2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017), pages 297–300. IEEE, 2017.
[40] Sonia Moreno-Grau, Octavio Rodriguez-Gómez, Angela Sanabria, Alba Pérez-Cordón, Domingo Sánchez-Ruiz, Carla Abdelnour, Sergi Valero, Isabel Hernandez, Maitée Rosende-Roca, Ana Mauleon, et al. Exploring apoe genotype effects on alzheimer’s disease risk and amyloid β burden in individuals with subjective cognitive decline: The fundacioAce healthy brain initiative (facehbi) study baseline results. Alzheimer’s & Dementia, 14(5):634–643, 2018. 3
[41] Basil Mustafa, Aaron Loh, Jan Freyberg, Patricia MacWilliams, Megan Wilson, Scott Mayer McKinney, Marcin Sieniek, Jim Winkens, Yuan Liu, Peggy Bui, et al. Supervised transfer learning at scale for medical imaging. arXiv preprint arXiv:2101.05913, 2021. 4
[42] Minh Nguyen, Tong He, Lijun An, Daniel C Alexander, Jiashi Feng, BT Thomas Yeo, Alzheimer’s Disease Neuroimaging Initiative, et al. Predicting alzheimer’s disease progression using deep recurrent neural networks. NeuroImage, 222:117203, 2020. 2
[43] Mehdi Noroozi and Paolo Favaro. Unsupervised learning of visual representations by solving jigsaw puzzles. In European conference on computer vision, pages 69–84. Springer, 2016. 1
[44] Kanghan Oh, Young-Chul Chung, Ko Woon Kim, Woo-Sung Kim, and II-Seok Oh. Classification and visualization of alzheimer’s disease using volumetric convolutional neural network and transfer learning. Scientific Reports, 9(1):1–16, 2019. 1
[45] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748, 2018. 1
[46] Ronald Carl Petersen, PS Aisen, Laurel A Beckett, MC Donohue, AC Gamst, Danielle J Harvey, CR Jack, WJ Ja gust, LM Shaw, AW Toga, et al. Alzheimer’s disease neuroimaging initiative (adni): clinical characterization. Neurology, 74(3):201–209, 2010. 3
Maithra Raghu, Chiyuan Zhang, Jon Kleinberg, and Samy Bengio. Transfusion: Understanding transfer learning for medical imaging. *Advances in neural information processing systems*, 32, 2019.

Shannon L Risacher, Andrew J Saykin, John D Wes, Li Shen, Hiram A Firpi, and Brenna C McDonald. Baseline mri predictors of conversion from mci to probable ad in the adni cohort. *Current Alzheimer Research*, 6(4):347–361, 2009.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

Udo Rueb, Katharina Stratmann, Helmut Heinsen, Kay Janani Venugopalan, Li Tong, Hamid Reza Hassanzadeh, of Alzheimer’s Disease Current Alzheimer Research cohort. dictors of conversion from mci to probable ad in the adni co.

Hiram A Firpi, and Brenna C McDonald. Baseline mri pre-

Bengio. Transfusion: Understanding transfer learning for computer vision and pattern recognition, pages 3733–3742, 2018.

Huidong Xie, Hongming Shan, Wenxiang Cong, Xiaohua Zhang, Shaohua Liu, Ruola Ning, and Ge Wang. Dual network architecture for few-view ct-trained on imagenet data and transferred for medical imaging. In *Developments in X-ray Tomography XII*, volume 11113, pages 184–194. SPIE, 2019.

Mang Ye, Xu Zhang, Pong C Yuen, and Shih-Fu Chang. Unsupervised embedding learning via invariant and spreading instance feature. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6210–6219, 2019.

Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stéphane Deny. Barlow twins: Self-supervised learning via redundancy reduction. In *International Conference on Machine Learning*, pages 12310–12320. PMLR, 2021.

Richard Zhang, Phillip Isola, and Alexei A Efros. Colorful image colorization. In *European conference on computer vision*, pages 649–666. Springer, 2016.

Hong-Yu Zhou, Shuang Yu, Cheng Bian, Yifan Hu, Kai Ma, and Yefeng Zheng. Comparing to learn: Surpassing imagenet pretraining on radiographs by comparing image representations. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 398–407. Springer, 2020.

Zongwei Zhou, Vatsal Sodha, Jiaxuan Pang, Michael B Gotway, and Jianming Liang. Models genesis. *Medical image analysis*, 67:101840, 2021.

Jiwen Zhu, Yuexiang Li, Yifan Hu, Kai Ma, S Kevin Zhou, and Yefeng Zheng. Rubik’s cube+: A self-supervised feature learning framework for 3d medical image analysis. *Medical image analysis*, 64:101746, 2020.

Xinrui Zhuang, Yuexiang Li, Yifan Hu, Kai Ma, Yijiu Yang, and Yefeng Zheng. Self-supervised feature learning for 3d medical images by playing a rubik’s cube. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 420–428. Springer, 2019.