ClinImCL: Self-Supervised Contrastive Learning for Longitudinal Clinical Imaging Analysis

Caden Roberts
cawrober@ucsc.edu

Sergio Casarrubias
secasarr@ucsc.edu

Aditya Sarin
adsarin@ucsc.edu

Abstract

For neuroimaging of Alzheimer's disease, the use of deep learning has led to a significant boost in diagnostic accuracy. However, in the hands of healthcare professionals, advanced deep neural network systems have remained somewhat underutilized. The sheer computational power and complexity of the models may hinder their ability to be deployed in clinical settings. When working with the OASIS-3 dataset, we employ the MONAI framework to efficiently process and learn representations of Alzheimer's disease progression from MRI scans, and present a self-supervised processing pipeline, ClinImCL, that efficiently learns representations from high-dimensional MRI data under constrained compute. In doing so, our goal is to meet the demands of methodological precision and real-world usability, making it possible to deploy deep learning models in resource-limited healthcare settings.

1 Introduction

In recent years, deep learning methods have emerged as powerful tools for predicting Alzheimer's disease, which has made a significant difference in how accurately we can diagnose Alzheimer's disease using brain MRI scans. With these advanced methods available, it has become much more reliable to detect early signs of Alzheimer's from images. Deep learning methods are extremely powerful, but the problem is that these methods and models require substantial computational power, such as high-end GPUs, large amounts of memory, advanced computational infrastructure, etc. Many clinics, and some hospitals, do not have access to these types of resources, which makes it difficult to deploy these models in areas with limited resources.

For example, a single brain magnetic resonance imaging scan can have millions of tiny, three-dimensional pixels called voxels, and each voxel captures detailed data on brain structures. To handle these data, deep neural networks are created, but they demand computing equipment. However, many hospitals and clinics lack these equipment, making it impossible for doctors to use these models.

Our research specifically tackles this issue. We used OASIS-3 LaMontagne et al. [2019], a publicly available data set that is a large collection of brain magnetic resonance images collected from more than 1300 people. An example is shown in Figure 1. This dataset contains 2842 MR sessions that cover healthy aging, as well as different stages of cognitive decline. We also used MONAI Cardoso et al. [2022] (Medical Open Network for AI), which is an open-source toolkit built on PyTorch, designed specifically to handle challenges associated with medical imaging, such as MRI scans. By combining OASIS-3 with MONAI's processing tools, we aim to streamline how we handle this large-scale medical imaging data and ultimately reduce the heavy computational demands associated with deep learning methods.

CSE 290D Neural Computation Project at UCSC "ClinImCL", NeurIPS 2025 AI4D3 2025 Workshop styling

In this research paper, we propose **ClinImCL**, self-supervised contrastive learning approach designed for longitudinal clinical imaging analysis. Our method leverages contrastive learning principles to capture changes in the brain's MRI scans over time, which reduces the need for large labeled datasets. ClinImCL aims to overcome equipment barriers in hospitals and clinics by efficiently learning from MRI data. We will evaluate our model's performance in its accuracy in detecting Alzheimer's disease progression. We also intend to show how our effective approach can reduce the computational demands efficiently, making it feasible for practical use in healthcare environments.

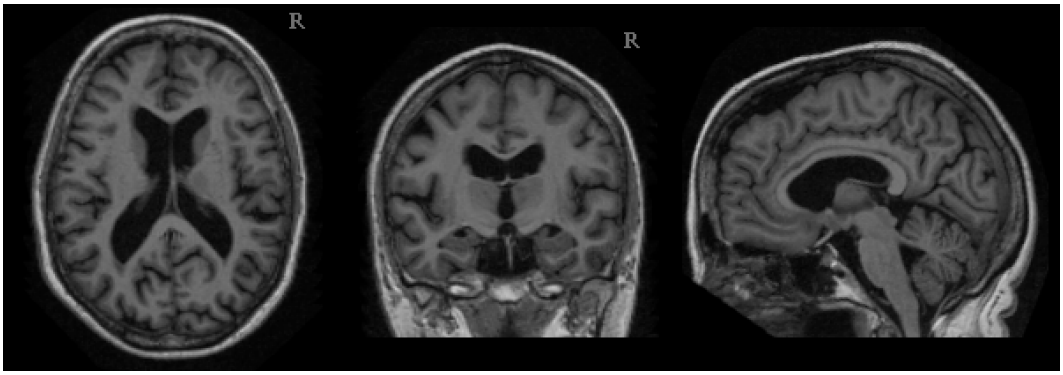

Figure 1: Example of OASIS-3 Scans

2 Methods

2.1 ClinImCL: Contrastive Learning Framework

Our proposed method, **ClinImCL**, adopts self-supervised contrastive learning tailored specifically for longitudinal clinical imaging analysis. The main objective is to learn robust, meaningful representations of brain MRI scans that efficiently capture subtle changes over time without relying on extensive labeled datasets.

ClinImCL operates in a single self-supervised representation learning stage, where embeddings are trained to align temporally related scans while separating scans from different subjects. Given MRI images x, we define positive pairs as scans from the same subject at different timepoints and negative pairs as scans from different subjects. The model is trained using the standard InfoNCE loss [van den Oord et al., 2019]:

$$\mathcal{L}_{\text{InfoNCE}} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{\exp(\text{sim}(z_i, z_i^+)/\tau)}{\sum_{j=1}^{N}\exp(\text{sim}(z_i, z_j)/\tau)}. \tag{1}$$

where sim denotes cosine similarity and $\tau = 0.07$ is the temperature parameter. This encourages the encoder to maximize similarity between longitudinally consistent representations while maintaining inter-subject separation.

2.2 Data Preprocessing

As a key step to ensure model efficiency, we applied a preprocessing pipeline specifically optimized for clinical MRI data. After initial downloading and verification, we utilized MONAI's comprehensive medical imaging tools for spatial normalization, intensity scaling, and foreground cropping. We resized the MRI scans to uniform spatial dimensions of 128^3 to ensure anatomical detail while maintaining efficiency.

2.3 Network Architecture

ClinImCL employs a modified custom lightweight 3D CNN architecture as the encoder network, adapted specifically for processing MRI data. The adaptation includes replacing standard two-

dimensional convolutions with three-dimensional convolutions, allowing it to capture volumetric information. The encoder provides an effective balance between complexity and computational efficiency, ideal for resource-limited clinical environments. The encoder produces embeddings which capture voxel representations of the input MRI scans.

2.4 Training and Validation Procedure

The training pipeline was implemented in Google Colab with NVIDIA A100 GPU acceleration, streaming all data directly from Google Cloud Storage to avoid local storage constraints. Preprocessed MRI tensors from the OASIS-3 dataset were accessed via `gcsfs`, normalized to the range [0,1], and resampled to a spatial resolution of 96^3 voxels for computational efficiency. Training pairs were generated dynamically, using scans from the same subject across visits as positives and scans from different subjects as negatives, allowing the model to capture temporal consistency and inter-subject variability.

A lightweight 3D convolutional encoder with four convolutional blocks and a projection head produced 128-dimensional normalized embeddings. The model was optimized with AdamW Loshchilov and Hutter [2019] using a learning rate and weight decay of 1×10^{-4}, and a temperature parameter of $\tau = 0.07$. Training was run for 10 epochs with automatic mixed precision, sampling approximately 25% of subjects per epoch to balance dataset coverage and runtime. Average epoch duration was 40 minutes. Model checkpoints were automatically saved and uploaded to Google Cloud Storage for persistent tracking.

Embedding quality was evaluated through UMAP [McInnes et al., 2020] projections of the learned 3D embeddings and via mean cosine distance between consecutive timepoints for each subject, which measures longitudinal stability. These confirmed that ClinImCL captured temporally consistent yet subject-specific trajectories, ensuring both computational efficiency and robustness in modeling longitudinal MRI progression from OASIS-3.

2.5 Computational Setup

Given the resource constraints common in clinical settings, a key focus of our research is computational feasibility. We evaluated ClinImCL under a realistic scenario of limited GPU availability, factoring computational costs including memory usage, training time, and inference speed in our Colab A100 High RAM environment. For data fetching and preprocessing, we used a Google Cloud VM with an n1-standard-8 and a disk size of 3000 GB. By systematically addressing computational efficiency, our approach aims to bridge the gap between advanced deep learning techniques and practical, clinical deployment scenarios.

3 Results

We evaluated the longitudinal contrastive model on 10 OASIS-3 subjects, each with multiple MRI timepoints, to assess whether embeddings captured within-subject temporal consistency while maintaining inter-subject separation. Across epochs, UMAP projections of the 3D embedding space revealed a clear convergence pattern. The InfoNCE objective decreased consistently from 2.05 in the first epoch to ≈ 1.94 by epoch 10, with a marked improvement between epochs 3–7 as the encoder's latent structure stabilized. Early training (epochs 1-3) produced disorganized trajectories with irregular transitions between scans, suggesting that the encoder had not yet learned stable temporal structure. By mid-training (epochs 5-7), subjects formed distinct and chronologically ordered trajectories, indicating that contrastive optimization had aligned embeddings along biologically meaningful temporal axes.

The embedding geometry reached its most coherent and interpretable form at epoch 7, shown in Figure 2. In this checkpoint, trajectories for all 10 subjects were smooth and well separated, showing minimal overlap while preserving progression continuity between visits. Corresponding cosine-distance metrics confirmed this visual trend: mean longitudinal distances were lowest and most uniform across subjects, reflecting maximal temporal stability without collapsing individual variation. This suggests that the model achieved an optimal balance between invariance (temporal alignment) and distinctiveness (subject identity).

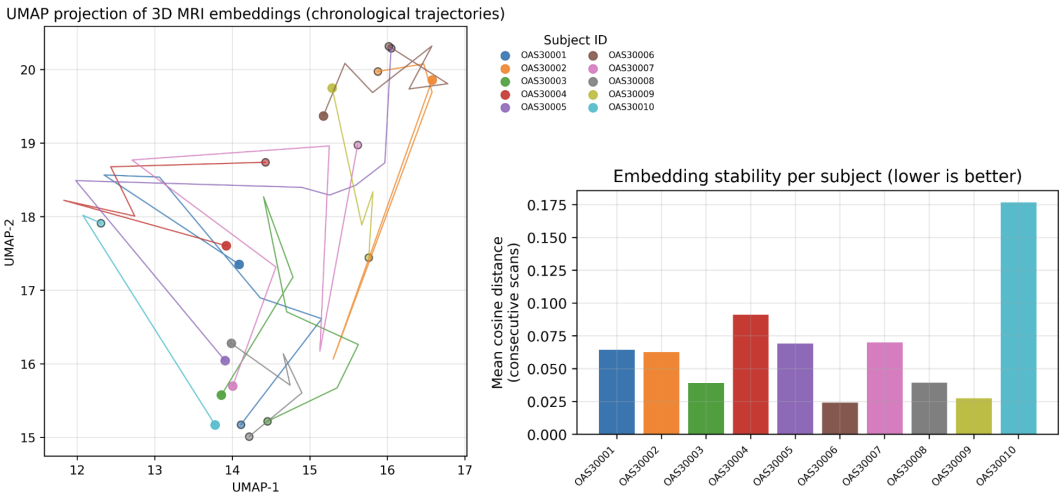

Figure 2: Epoch 7

After epoch 7, further training (epochs 8-10) seemed to introduce over-compression, where trajectories became overly linear and global variance decreased. Although the overall loss continued to fall, UMAP topology indicated reduced subject-specific nuance and increasing trajectory alignment along a dominant axis, which are hallmarks of representational overfitting. Thus, performance stabilized and then slightly deteriorated beyond the seventh epoch, marking it as the optimal checkpoint for downstream use and longitudinal interpretation.

4 Discussion

These embeddings may provide a strong foundation for future analysis of temporal consistency and disease progression, but significant model results are pending greater epoch completions. Initial training on Google Cloud was constrained by limited GPU access. Switching to a Colab Pro A100 instance enabled stable mixed-precision runs, though streaming data from cloud storage introduced additional overhead. Our ongoing work will focus on optimizing the contrastive learning framework, reducing computational overhead, and evaluating how effectively the learned embeddings can support clinically relevant tasks such as Alzheimer's disease classification and progression estimation. Future work could extend ClinImCL to larger datasets and evaluate its clinical transferability.

References

M. Jorge Cardoso, Wenqi Li, Richard Brown, Nic Ma, Eric Kerfoot, Yiheng Wang, Benjamin Murrey, Andriy Myronenko, Can Zhao, Dong Yang, Vishwesh Nath, Yufan He, Ziyue Xu, Ali Hatamizadeh, Andriy Myronenko, Wentao Zhu, Yun Liu, Mingxin Zheng, Yucheng Tang, Isaac Yang, Michael Zephyr, Behrooz Hashemian, Sachidanand Alle, Mohammad Zalbagi Darestani, Charlie Budd, Marc Modat, Tom Vercauteren, Guotai Wang, Yiwen Li, Yipeng Hu, Yunguan Fu, Benjamin Gorman, Hans Johnson, Brad Genereaux, Barbaros S. Erdal, Vikash Gupta, Andres Diaz-Pinto, Andre Dourson, Lena Maier-Hein, Paul F. Jaeger, Michael Baumgartner, Jayashree Kalpathy-Cramer, Mona Flores, Justin Kirby, Lee A. D. Cooper, Holger R. Roth, Daguang Xu, David Bericat, Ralf Floca, S. Kevin Zhou, Haris Shuaib, Keyvan Farahani, Klaus H. Maier-Hein, Stephen Aylward, Prerna Dogra, Sebastien Ourselin, and Andrew Feng. Monai: An open-source framework for deep learning in healthcare, 2022. URL `https://arxiv.org/abs/2211.02701`.

Pamela J. LaMontagne, Tammie LS. Benzinger, John C. Morris, Sarah Keefe, Russ Hornbeck, Chengjie Xiong, Elizabeth Grant, Jason Hassenstab, Krista Moulder, Andrei G. Vlassenko, Marcus E. Raichle, Carlos Cruchaga, and Daniel Marcus. Oasis-3: Longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer disease. *medRxiv*, 2019. doi: 10.1101/2019.12.13.19014902. URL `https://www.medrxiv.org/content/early/2019/12/15/2019.12.13.19014902`.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization, 2019. URL https://arxiv.org/abs/1711.05101.

Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction, 2020. URL https://arxiv.org/abs/1802.03426.

Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding, 2019. URL https://arxiv.org/abs/1807.03748.