

Imaging-based prediction of transcriptional subtypes in Alzheimer’s disease

Tony Ding

TDING@MIT.EDU

Harvard T.H. Chan School of Public Health

Benjamin James

BENJAMES@MIT.EDU

MIT CSAIL

Broad Institute of MIT and Harvard

Guangze Luo

GUANGZEL@MIT.EDU

Harvard T.H. Chan School of Public Health

Ananth Shyamal

ASHYAMAL@MIT.EDU

MIT Department of Biology

MIT Department of Electrical Engineering and Computer Science

Mathilde Tans

MTANS@MIT.EDU

Harvard T.H. Chan School of Public Health

Ryan Wang

RYANWANG@MIT.EDU

Harvard T.H. Chan School of Public Health

Harvard Department of Biostatistics

Boshen Yan

BOSHENYA@MIT.EDU

Harvard Medical School

Chen Yang

CHENY357@MIT.EDU

Harvard T.H. Chan School of Public Health

Abstract

This study explores the application of genetic information, and deep learning techniques to predict Alzheimer’s Disease (AD). For brain MRI image analysis, we compared the results using Convolution Neural Networks (CNN) built from scratch, CNN transfer learning, and Vision Transformers (ViT). Our vision transformer model achieved the best performance on the test set with 98.57% accuracy. We explored the method using ViT in 3D brain MRI image analysis by treating each 3D patch as a vector to feed into the transformer. Using CNN/ViT model, we also extracted feature representations of brain MRI images for multi-modal usage. By combining feature encoders from image and genetic deep learning models, we are able to further improve performance using a self-supervised contrastive learning pipeline. We show that feature explanation techniques can attribute genetic contributions to image embeddings, even after fine-tuning, which can be effective in understanding the genetic impact of neuro-imaging in Alzheimer’s Disease. For the first time to our knowledge, we demonstrated the feasibility of cell type specific PGS and their improved performance in (1) predicting white blood cell count in PBMCs, as well as a more complex phenotype, gray matter volume in the frontal cortex.

1. Introduction

Alzheimer’s disease (AD) is a complex neurodegenerative disease that affects millions of people worldwide. Its heterogeneity in phenotype, pathology, and cognition makes it challenging to diagnose and treat. Recent advancements in single-cell genomics and large-scale genetic studies provide an opportunity to understand the molecular basis underlying phenotypic heterogeneity in AD. Our project aims to leverage deep learning techniques to analyze brain MRI images, patient omics datasets, as well as demographics and clinical information in the hopes of identifying patterns and associations between brain-MRI traits, transcriptional subtypes, pathology, and cognition in AD. This can potentially lead to earlier detection, better patient outcomes, and a deeper understanding of the disease’s mechanisms.

Generalizable Insights about Machine Learning in the Context of Healthcare

In this project, we explored the application of various methodologies to analyze brain MRI images and patient omics to predict phenotypes related to Alzheimer’s disease. Specifically, we explored:

1. The use of cell type-specific polygenic scores to predict brain-related traits
2. Neuro-imaging using CNNs, transfer learning, and vision transformers to predict Alzheimer’s disease
3. The use of self-supervised contrastive learning to combine polygenic and neuro-imaging to learn joint explainable representations

1.1. Cell Type-specific Polygenic Scores

Polygenic scores (PGS) have revolutionized clinical data, providing insight orthogonal to electronic health records and family history. With the advent of large-scale genetic studies, including millions of disease-SNP associations in the GWAS Catalog, and the ever-growing genetic cohorts like the UK Biobank and FinnGen, genetics is quickly transforming the field. The PGS Catalog currently has 465 publications contributing 3640 PGS to its database. However, since PGS have not performed better than a simple linear model, the mechanistic insights of PGS have not been included.

1.2. Deep Learning for Neuro-imaging

Convolutional Neural Networks (CNNs) represent a category of deep learning algorithms, particularly renowned for their efficacy in image recognition tasks [Krizhevsky et al. \(2017\)](#). One of the key strengths of CNNs lies in their capacity to learn spatial hierarchies of features in an automatic and adaptive manner, which is extremely beneficial for image analysis [Le-Cun et al. \(2015\)](#). Therefore, it is our first choice to analyze brain MRI images and serves as a baseline model to offer an intuitive starting point. Nevertheless, CNNs are not innately invariant to scale and rotation and can encounter difficulties when such variations are present in the data. In addition, their requirement for large volumes of data and substantial computational resources for training can pose challenges. This is where the concept

of transfer learning becomes particularly appealing.

Transfer learning stands as an efficient approach to achieve high performance with fewer data than typically required for deep learning models, thereby conserving considerable computational resources, given that a portion of the model has already been trained [Yosinski et al. \(2014\)](#). VGG16 model [Simonyan and Zisserman \(2014\)](#), a deep CNN trained on ImageNet, is particularly appealing for our task due to its proven capability to capture complex patterns in images and its relative simplicity, which helps prevent overfitting. Nevertheless, transfer learning can still be limited by significant discrepancies between datasets and potential overfitting issues.

Instead, the Vision Transformers (ViT) model is not primarily built on the convolution operation and it leverages the strengths of the transformer architecture, which has been instrumental in achieving state-of-the-art performance in Natural Language Processing (NLP) tasks. [Dosovitskiy et al. \(2020\)](#).

ViT is composed of a series of transformer blocks; these blocks feature a self-attention mechanism, which allows the model to focus on different parts of the input image, treating them as a sequence of patches. This mechanism enables the model to capture both local and global image features, a capability that can be vital in complex tasks such as our analysis of MRI images for Alzheimer’s disease.

Additionally, the ViT model includes a patch embedding layer that disassembles the input image into smaller patches and projects each patch into a vector linearly. This characteristic enables the ViT model to accommodate variable-sized input images. Their superior performance in capturing global patterns in an image makes them particularly suitable for complex tasks such as 3D MRI imaging data. Given the best performance in 2D brain MRI imaging data in our final result, ViT would be our potential first choice if we could have obtained 3D imaging data.

1.3. Understanding Cross-modal Behavior in AD

AD is a multi-faceted disease associated with various biomarkers and expression through multiple modalities, both genetic, socio-economic, and neuro-imaging [Knopman et al. \(2021\)](#). As such, understanding the impact each modality has on one another can help deeper understand disease interaction. Furthermore, this can further improve diagnostic tools by combining information contributed by various modalities. We aim to investigate the connection between genotypic and lab modalities using traditional models (logistic regression, tree-based methods, etc.), and the connection between genotypic and imaging modalities through ContIG, a more explainable self-supervised contrastive deep learning framework [Taleb et al. \(2022\)](#).

Self-supervised contrastive learning separates dissimilar observations from similar ones without labels, learning more task-agnostic embeddings and can be fine-tuned for applications such as image classification with high performance [Chen et al. \(2020\)](#). The previously men-

tioned framework, ContIG, extends this to multiple modalities and allows for cross-modal feature explanations. As this method shows superior performance in both prediction and explainability, we implement it to investigate the complex interactions of genetic-neuro-imaging modalities and multi-modality-informed prediction.

2. Related Work

2.1. Genetics

Polygenic scores have been deposited to the PGS Catalog at a large rate. Traditionally, genetic effects are understood linearly, as a sparse Lasso-based regression problem. Previously, different variants have been understood based on basic genomic information, such as synonymous or non-synonymous exon variant, intron, promoter variant, UTR variant, or intergenic variant should have different effect sizes. In a previous study, that was precisely done, modifying regularization to allow for different Lasso regularization [Tanigawa et al. \(2022\)](#). Lasso-based methods such as SNPNET [Li et al. \(2022\)](#) allow for large-scale genetic tests for disease, using millions of SNPs from the UK Biobank. We plan to include epigenomic information to further model disease in each cell type. To our knowledge, cell type-specific PGS has not been explored yet.

2.2. Vision Models

Research on leveraging vision models, particularly deep learning architectures, for Alzheimer’s Disease diagnosis has been a rapidly evolving field in recent years. Several studies have demonstrated the potential of deep learning techniques for extracting meaningful features from neuroimaging data, aiding in the diagnosis of AD. [Sarraf and Tofghi \(2016\)](#) applied a Convolutional Neural Network model to classify AD patients using MRI and fMRI data, obtaining promising results. [Chabib et al. \(2023\)](#) introduced a curvelet transform-based-convolutional neural network to detect early-stage AD using MRI images, and [Folego et al. \(2020\)](#) constructed a 3D-CNN model to classify structural MRI images into various levels of cognitive impairment, both obtaining competitive results.

Additionally, there has been a recent surge in research exploring transformer-based architectures for medical imaging, including AD. [Chen et al. \(2021\)](#) proposed TransUNet, a hybrid transformer, and UNet model, demonstrating its effectiveness on various medical imaging tasks. [Sarraf et al. \(2023\)](#) introduced an optimized vision transformer architecture to predict different stages of AD using MRI images. [Dhinagar et al. \(2023\)](#) applied several vision transformer variants to 3D brain MRI images, experimenting with various techniques that are essential for training vision transformer-like models.

Lastly, research on transfer learning in the context of AD imaging has also gained momentum. [Hosseini-Asl et al. \(2016\)](#) utilized a transfer learning approach, implementing a pre-trained CNN model for AD diagnosis. Their model, trained on a large-scale natural image dataset, was fine-tuned on a smaller AD dataset, showcasing its power in capturing the intricate patterns in brain MRI images. [Acharya et al. \(2021\)](#) used multiple trans-

fer learning models to predict multi-class AD, presenting efficacy and improved accuracy. [Sharma et al. \(2022\)](#) used the VGG16 model for early prediction of AD.

2.3. Deep Learning for Multiple Modalities

Multi-modal learning has been explored for Alzheimer’s disease prediction via methods such as the concatenation or aggregation of features from multi-modal encoders, where the resulting aggregated feature is used in a downstream classifier [Venugopalan et al. \(2021\)](#), [Wang et al. \(2022\)](#). In other domains, such as visual-textual domains, contrastive methods have been shown to outperform state-of-the-art [Radford et al. \(2021\)](#), [Khosla et al. \(2021\)](#). Furthermore, these methods have been applied to general areas of the medical field with success [Taleb et al. \(2022\)](#), [Zhang et al. \(2020\)](#). In the context of Alzheimer’s Disease, contrastive methods have been applied [Jung et al. \(2022\)](#), however deep exploration into multi-modal contributions to AD using these methods remains an open question, both in terms of predictive performance and explainability.

3. Methods

3.1. Cell Type-specific PGS for Brain Related Traits

3.1.1. WEIGHTING OF REGULATORY ELEMENTS USING PBMCs IN EpiMAP TO PREDICT WHITE BLOOD CELL COUNT

We want to weight regulatory elements, specifically enhancers, and promoters in the relevant cell type(s), more in a PGS to provide a more mechanistic insight into PGS, making a giant regression not only more interpretable but also likely more relevant clinically. Using SNPNET [Li et al. \(2022\)](#), instead of restricting SNPs to fall inside PBMC enhancers or promoters, we determined that modifying the Lasso weights (the p.factor parameter in SNPNET) would lead to a better result since we want these PGS to pass the un-weighted PGS.

3.1.2. PREDICTION OF MRI-BASED PHENOTYPES USING FRONTAL CORTEX AND NUCLEUS ACCUMBENS SNATAC-SEQ DATA

Cell type-specific regulation has moved to single-cell ATAC sequencing. Moreover, large atlases have been built. Currently, in the Kellis lab, multi-omic snRNA+snATAC-seq has been sequenced, and Frontal cortex (BA9-FC) and nucleus accumbens (NAc) samples have been acquired. We use the accessible sites per cell type in these two brain regions to similarly guide PGS to be cell type-specific. Moreover, we use the 0.85 weighting in SNPNET used in the PBMC/EpiMap PGS for white blood cell count. Next, for the MRI-based phenotypes, we used gray matter volume traits in the UK Biobank. In over 40,000 individuals, this MRI data is available and measured the R2 value between the PGS and the observed trait. For these predictions, we utilized the following covariates: 10 genetic principal components, age, sex, and Townsend deprivation index, which is a socioeconomic status indicator.

3.2. Gene-set Analysis of GWAS Data Using MAGMA

3.2.1. GWAS BACKGROUND

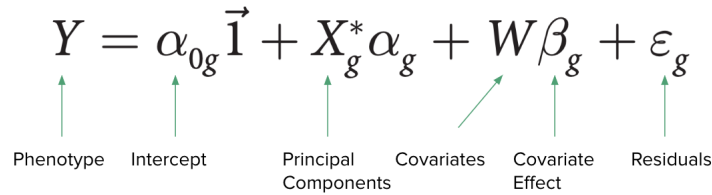
A genome-wide association study (GWAS) is a common method to determine what genes, and in particular, what single nucleotide polymorphisms (SNPs), contribute to disease (phenotype of interest). In a GWAS, the genomes of people exhibiting the phenotype (patient group) are compared to the genomes of people who do not exhibit the phenotype (control group). Specifically, differences in SNP occurrence are observed between the two groups. Then, we can associate SNPs to regions of chromosomes that may cause disease and determine SNP p-values.

GWAS is a powerful diagnostic tool to determine patients who are predisposed to certain diseases, e.g., neurological diseases. Additionally, GWAS can give us a broad idea, at the single chromosome level, as to where disease-causing genes may reside. However, GWAS cannot identify specific genes involved in diseases. Furthermore, GWAS are severely limited by population size, as many diseases are highly polygenic, having thousands of variants with very small effects on the phenotype.

3.2.2. GENE SET ANALYSIS AND MAGMA

The shortcomings of GWAS have led to the advent of gene analysis and gene-set analysis methods by aggregating the SNP data in GWAS to the whole gene level or groups of genes sharing similar characteristics. Genetic markers in the same gene or gene set are jointly associated, increasing the ability to find significant phenotypic effects amongst markers with individually weak associations. There are several existing methods of gene and gene-set analysis, but it turns out that the statistical power of most of these methods is limited by linkage disequilibrium.

One method that has shown superior statistical power and speed is MAGMA (Multi-marker Analysis of GenoMic Annotation) [Christiaan A. de Leeuw \(2015\)](#). MAGMA uses a

$$Y = \alpha_{0g} \vec{1} + X_g^* \alpha_g + W \beta_g + \varepsilon_g$$


The diagram shows the equation $Y = \alpha_{0g} \vec{1} + X_g^* \alpha_g + W \beta_g + \varepsilon_g$ with green arrows pointing from labels below to terms in the equation:

- Y is labeled "Phenotype"
- α_{0g} is labeled "Intercept"
- X_g^* is labeled "Principal Components"
- α_g is labeled "Covariates"
- W is labeled "Covariate Effect"
- β_g is labeled "Covariate Effect"
- ε_g is labeled "Residuals"

Note: The arrow for α_g points to the label "Covariates", and the arrow for W points to the label "Covariate Effect".

Figure 1: MAGMA utilizes a multiple linear regression model for gene analysis.

multiple linear principal components regression model to account for LD when performing gene analysis. MAGMA works by first isolating the principal components of the SNP matrix from GWAS data, discarding the ones that have relatively small eigenvalues, and then feeding this along with a covariates matrix into the regression model. Overall, the well-documented performance of MAGMA on GWAS datasets and its convenient software download made it a suitable tool for us to conduct gene-set analysis.

3.3. Brain MRI Image Analysis

3.3.1. CONVOLUTIONAL NEURAL NETWORK

A CNN architecture was designed for our classification task. The architecture consists of several convolutional layers, followed by pooling layers, drop-out layers, fully connected layers, and a final output layer. Each convolutional layer applies a series of filters to the input data, capturing local features of the image, such as edges, textures, and shapes. The output of each layer is typically passed through ReLU. It was typically trained using back-propagation and a variant of gradient descent and categorical cross-entropy loss function is used to detect Alzheimer’s.

3.3.2. CNN TRANSFER LEARNING

Given the limited size of our medical imaging dataset, transfer learning was employed to ameliorate the performance of our model. The VGG16 model, pre-trained on the ImageNet dataset, was chosen for this task due to its robust performance in image recognition. The initial layers of VGG16, which capture generic features, were frozen during training, while the classification layers that can capture more complex and task-specific features were fine-tuned based on our dataset.

3.3.3. VISION TRANSFORMER

Vision transformer is a technique that has shown superior performances on many imaging tasks. One of the main advantages of ViT is its ability to capture long-range dependencies in the image. Unlike CNNs, which rely on local feature extraction through convolutions, ViT uses a self-attention mechanism that allows it to attend to the entire image at once, thus capturing global information and important relationships between different parts of the image. In our case, we applied the vision transformer: vit_b32 (i.e. this ViT model has 32 transformer blocks) from the vit.keras package in order to leverage the benefits of the transformer architecture. We also added several dense layers, batch normalization layers, and dropout layers for image classification purposes. Finally, we intended to train our ViT model for over 1000 epochs with a low learning rate of 0.0001, a corresponding decay rate of 0.00001, and an early stopping mechanism with a patience of 50. All code for the image analyses can be found at <https://github.com/MathildeTans/MLHC-Project-Imaging>.

3.4. Multimodal Analysis of Alzheimer’s Disease

In order to further improve imaging performance, we aim to pre-train genetics-informed image embeddings. We believe this will be especially helpful in the case where our target task has a small amount of labeled data (e.g. target AD subtype data), while we have a large amount of unlabeled multi-modal data (e.g. UK Biobank). We also aim to evaluate this model’s performance on genetic-demographic-clinical data. Code can be found at <https://github.com/rjunw/ad-multimodal/>.

3.4.1. SELF-SUPERVISED CONTRASTIVE LEARNING

We follow the self-supervised contrastive learning pipeline, ContIG, as described in Taleb et al. (2022). This setting supposes we have a multimodal dataset $\mathcal{D} = \{x_i, g_i, y_i\}_{i=1}^N$, where

an individual i has some image x_i , some genetic data g_i , and some label y_i . We use the following bi-directional cross-modal contrastive loss as defined by Taleb et al. (2022)

$$\mathcal{L}(x, g) = \lambda L(x, g) + (1 - \lambda) L(g, x) \text{ where } L(x, g) = - \sum_{i=1}^b \log \frac{\exp\{\cos(z_i^x, z_i^g)/\tau\}}{\sum_{k=1, k \neq i}^b \exp\{\cos(z_i^x, z_k^g)\}}$$

$L(x_i, g_i)$ amounts to holding x_i ’s embedding z_i^x fixed and pulling genetic embedding, z_i^g closer. The rest of the genetic embeddings in batch of size b , $\{z_k^g\}_{k=1, k \neq i}^b$, are pushed apart.

3.4.2. IMAGE-GENETICS MODEL ARCHITECTURE

Due to its low computational requirements, we use a ResNet-18 He et al. (2015) feature encoder with an MLP projection head; however, the feature encoder can be replaced with any model in §3.3. The projection head has one 512-unit layer and outputs a 64-dimensional latent vector z_i^x . The genetic encoder is given by an MLP, taking a genetic sequence input to a 64-dimensional latent vector z_i^g . For classification, we consider image inputs and discard the genetic encoder and projection head as in Chen et al. (2020). We append a two-layer MLP to the ResNet-18 feature encoder and fine-tune the new classifier using cross-entropy loss. We use a batch size of 128. We compare this to a baseline non-pre-trained ResNet-18 with the same classification head. We use the same hyperparameters for both classification models in order to compare the best model performance based on the best validation loss.

3.4.3. GENETIC EXPLANATIONS OF IMAGES

We use Integrated Gradients to explore the feature attribution of genetics data on image embeddings Sundararajan et al. (2017), Taleb et al. (2022), Kokhlikyan et al. (2020). Let $x_{(i)}$ be the i^{th} dimension of input x , x' be a baseline input, and f be a model.

$$\text{IntegratedGrads}_i(x) := (x_{(i)} - x'_{(i)}) \times \int_{\alpha=0}^1 \frac{\partial f(x' + \alpha \times (x - x'))}{\partial x_{(i)}} d\alpha$$

This can be interpreted as interpolating gradients along the baseline (without input feature) to the input, resulting in the feature contribution of dimension i . In accordance to Taleb et al. (2022), we define a fixed reference $R = \{x_r, g_r\}_{r=1}^{512}$ batch for contrastive loss. We then evaluate contributions of individual i ’s genetic data g_i on image x_i by using Integrated Gradients over the loss function defined by $E(x_i, g_i) := \mathcal{L}(R \cup \{x_i, g_i\})$.

3.4.4. ANALYSIS OF GENOTYPE-CLINICAL-DEMOGRAPHIC MODALITIES

We trained and fine-tuned several models, including elastic net logistic regression, balanced random forest, gradient boosting trees, and a 4-layer multi-layer perceptron (MLP) with hidden layers of size 50, 25, 10. To account for data imbalance, class weights were set to be inversely proportional to the frequency of each class. Model performances were benchmarked with a majority class classifier. In addition, we trained a 30-dimensional joint embedding of genotype features and clinical/demographic features using the ContIG framework, which is attached to the same MLP framework for prediction. Model performances were evaluated based on accuracy, the area under the receiver operator curve (AUC), as well as sensitivity, and specificity.

4. Data and Experiment Setup

4.1. UK Biobank

Our dataset consists of a large cohort of about 500,000 individuals from the United Kingdom who participated in the UK Biobank study. For our study, we extracted 500 patient single nucleotide polymorphism (SNP) genotypes with known associations with AD (EBI trait ID MONDO_0004975). We also extracted participants’ detailed phenotypes, including blood and urine blood markers, ICD-10 codes and top 40 principal components of participants’ genotypes. Data is obtained from the Luria cluster, while data access is governed under the data usage agreement between the Kellis lab and UK Biobank under Application Number 21942, “Integrated models of complex traits in the UK Biobank”. For the polygenic scores, we utilized the trait f.25831.2.0 for the MRI based analysis, corresponding to Volume of gray matter in frontal medial cortex (right). For all PGS analyses, we included 10 PCs, age, sex, and Townsend deprivation index as covariates, used only the White British (WB) individuals, and used the train-test-validation split used in the Kellis lab, which is balanced among meaningful covariates and should not leak information.

4.1.1. JOINING DATA MODALITIES

While we had initially hoped to obtain a multimodal dataset consisting of patient genotypes, phenotypes, and image modalities, the data available to our team was badly harmonized. Despite our best efforts, we were not able to obtain access to brain MRI images for analysis. In addition, there is a mismatch between the participant IDs used in the genotype and phenotype datasets, which prevented us from performing joint analysis.

4.1.2. PHENOTYPE DATASET

Our extracted phenotype dataset (n=502460) comprises 50 anthropometry, blood, and urine biomarker features and the top 40 principal components of participants’ genotypes. In addition, participants’ family history of AD was labeled based on the self-reported occurrence of AD in their parents’ or siblings’ history of illnesses. Participants were labeled as AD positive if they had been diagnosed with one or more of the following AD-related ICD-10 codes: G300, G301, G308, G309, F000, F001, F002, F009. Based on this definition, we have identified 2765 AD patients in the cohorts, representing 0.55% of the cohort.

As the dataset is extremely imbalanced and that the feature dimension is low, we had opted for a random downsampling of the non-AD cohort from 499,696 to 50,000. For our analysis, we excluded 7 phenotype features with more than 30% missingness. The dataset was then split into training, validation and testing sets with a 70:10:20 split while keeping the proportion of AD patients in each split approximately equal. To avoid data leakage, missing values were filled with the mean value of the training set. Phenotypic features were then scaled using Scikit-learn’s implementation of StandardScaler using the feature value distribution of the training set. Genotype principal components were not re-scaled.

4.2. Brain MRI Image Analysis - Dataset from Kaggle

In direct imaging analysis, we utilized the Alzheimer MRI Preprocessed Dataset, available on Kaggle [ADNI \(2004, 2011\)](#); [OASIS \(2007, 2010\)](#); [Boysen \(2017\)](#); [Data.gov \(2023\)](#); [CORDIS \(2021\)](#); [Kumar and Shastri \(2022\)](#). The dataset comprises preprocessed Magnetic Resonance Imaging (MRI) scans from various sources, including public repositories, hospitals, and websites. The data was gathered with the intent to facilitate the design and development of accurate frameworks or architectures for Alzheimer’s Disease classification.

The dataset contains a total of 6400 MRI images, all of which have been resized to a uniform resolution of 128 x 128 pixels. These images are categorized into four distinct classes based on the severity of dementia: Mild Demented (896 images), Moderate Demented (64 images), Non-Demented (3200 images), and Very Mild Demented (2240 images)

4.3. Simulated Multimodal Data

4.3.1. NON-SIMULATED IMAGING DATA

In order to verify the effectiveness of the self-supervised contrastive learning framework from [Taleb et al. \(2022\)](#) on image-genetic modalities, we use publicly available MRI imaging data [Bhuvaji \(2020\)](#). These data include 2870 training examples in four classes, glioma tumor, meningioma tumor, pituitary tumor, or no tumor. We split these into train, validation, and test splits using 70%, 15%, and 15% of the data, respectively. We train our baseline ResNet-18 solely on these image data. We pre-train our contrastive ResNet-18 on this imaging data as well as simulated genetic data, followed by further fine-tuning on these image data. These data and labels serve as a proxy for images that will be further applied once UK Biobank and AD transcriptional subtype data has been made available.

4.3.2. SIMULATED GENETIC DATA

We simulate genetic data by correlating the number of non-reference alleles (0, 1, or 2) with image labels, mimicking raw SNP data. We simulate a total of 500 SNPs per individual. A label of $y_i \in \{0, 1, 2, 3\}$ will correspond to SNPs $y_i \times 50$ to $(y_i + 1) \times 50$ sampling with $p_0 = 0.05, p_1 = 0.05$, and $p_2 = 0.90$ for non-reference alleles 0, 1, or 2, respectively. Here $y_i = 0$ is a glioma tumor, $y_i = 1$ is a meningioma tumor, $y_i = 2$ is no tumor, and $y_i = 3$ is a pituitary tumor. A higher p_2 implies two copies of the non-reference allele will be more associated with the given label. All other SNPs for a given individual will be simulated with $p_0 = 0.25, p_1 = 0.55$, and $p_2 = 0.20$ for a more balanced SNP set. For example, the first 50 simulated SNPs in a $y_i = 0$ individual are more likely to be 2 than 0 or 1.

5. Results

5.1. Genetic Analysis

5.1.1. CELL TYPE-SPECIFIC PGS ON EPIMAP PBMCs

In order to determine the feasibility of using cell type-specific epigenomic signal to more relevantly partition PGS, we had to use a well-powered example for a strong genetic trait with very well-annotated data. Blood traits are very well annotated in the UK Biobank

with strong genetic heritability, and GWAS for these traits are well-powered. Additionally, blood (PBMCs) are extremely well profiled due to their clinical significance as a marker. In EpiMap Boix et al. (2021), many blood samples were included. We used biosample BSS01279 (PBMCs from ENCODE) since it was the most recently generated PBMC sample among 6 PBMC samples, out of 40 well-annotated blood and T cell biosamples. We trained SNPNET Li et al. (2022) on enhancers and promoters in PBMCs. We utilized the first 50 iterations (out of 100) due to time and computational constraints to gauge the model. Thus, R^2 values are not very high; as seen in the figure below, R^2 is gradually increasing with more iterations.

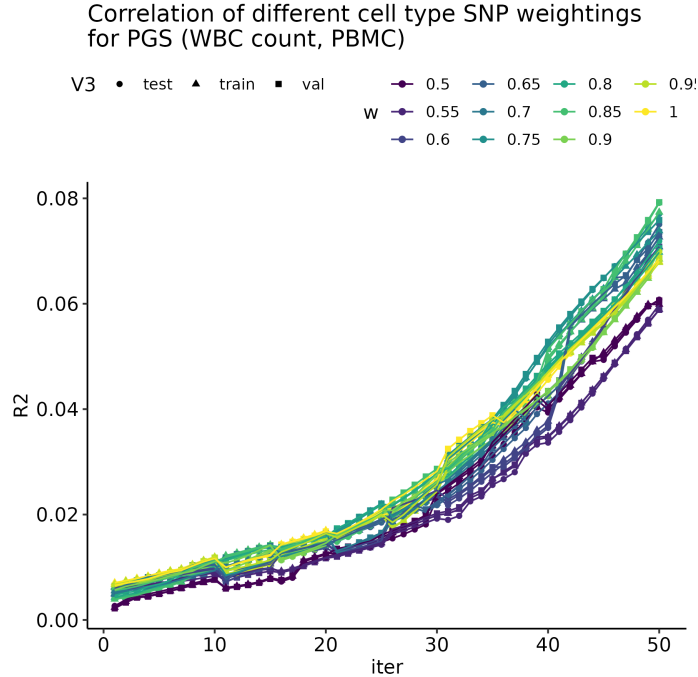


Figure 2: Weightings of enhancers in SNPNET model. Training, test, and validation Spearman R^2 between the PGS and the white blood cell count (WBC) on the PBMC data was used to judge. In the regime of 35-50 iterations, with the most terms in the Lasso model, the weighting of 0.85 seems to generate the highest weighting. The baseline weighting of 1 is among the lowest R^2 . This result confirms that cell type specific PGS indeed works and generates more relevant phenotypes than the raw models.

5.1.2. CELL TYPE SPECIFIC PGS ON BRAIN MRI TRAITS

To connect genetic variants to Brain MRI data, we utilized parameters learned from the PBMC analysis, namely a weight of 0.85 for accessible peaks in each cell type. Accessible peaks were gathered from snATAC-seq data from prefrontal cortex (BA9-PFC) and nucleus accumbens (NAc), provided by Benjamin from his research. Similarly seen in the PBMC re-

sults, the cell type specific PGS again more accurately predicted the trait of interest (Brain volume). Interestingly, the PGS with the most correlation with frontal cortex volume is Inhibitory neurons (overall), and deeper layer excitatory neurons, e.g. L6b and L5-6, had higher correlations to the gray matter volume than the lower layer excitatory neurons.

When looking at the top traits, we find that the GREAT McLean et al. (2010) enrichments of the peaks indicates cell type specificity of the cell types implicated, meaning that the cell types of interest are implicated in this model.

Furthermore, when all cell type-specific PGS are combined in an ensemble model using linear regression, a test R^2 of 0.7335 is achieved, compared to a test $R^2=0.7073$ in the highest non-ensemble model. This demonstrates that combining multiple PGS can lead to a greater outcome.

5.1.3. GENE-SET ANALYSIS OF BRAIN VOLUME MEASUREMENT GWAS DATA

We used GWAS catalog data from the National Human Genome Research Institute, and one of the trait categories we investigated was "brain volume measurement" (EFO_0006930). We parsed through all 3865 associations across 1125 studies and isolated relevant SNP location and GWAS p-value information.

We used MAGMA to run gene annotation to map the SNPs in the dataset to genes, perform gene analysis to compute p-values using a reference dataset to account for linkage disequilibrium ("g_1000_eur" from the 1000 Genomes Project), and perform gene-set analysis. In the gene-set analysis step, we used gene sets based on the AD transcriptional model proposed by Tanigawa (2023). The code can be found at <https://github.com/ashyamal/MLHC>.

The results of the MAGMA gene-set analysis are shown in Figure 3(next page). The high p-values indicate the insignificance of the gene sets with respect to brain volume measurement GWAS data. However, transcriptional subtype 14 showed much higher association (lowest p-value) compared to other subtypes, with subtypes 9, 11, and 17 also showing improved association. These results indicate that while the brain volume measurement data exhibited some level of association with AD, improved stratification into brain-specific regions is necessary.

5.2. Brain MRI Image Analysis

5.2.1. CNN

Upon completion of the training process, we arrived at a final training loss for CNN of 0.0294 and an impressive training accuracy of 0.9919. The performance on the validation dataset was also robust, with a validation loss of 0.0862 and a validation accuracy of 0.9703. As depicted in the accompanying figures, the training accuracy demonstrated a consistent upward trajectory, stabilizing around the 15th epoch. Similarly, the training loss steadily decreased until the 28th epoch, beyond which it plateaued. This plateau suggests that the model had effectively learned the salient features from the training data, hence the insignif-

VARIABLE	NGENES	BETA	BETA-STD	SE	P
Tx1	17	-0.035667	-0.0057486	0.3703	0.53835
Tx2	29	-0.028662	-0.005975	0.29924	0.53814
Tx3	2	-0.098749	-0.0055246	1.0821	0.53634
Tx4	23	-0.0018653	-0.00034799	0.32503	0.50229
Tx5	10	0.56288	0.06997	0.41318	0.086805
Tx6	15	-0.8003	-0.12136	0.37344	0.98375
Tx7	5	-1.9705	-0.17389	0.58833	0.99957
Tx8	12	-0.54851	-0.074573	0.38765	0.9212
Tx9	9	-0.24254	-0.028625	0.49264	0.68866
Tx10	29	0.23256	0.04848	0.29451	0.21502
Tx11	19	0.50023	0.085097	0.34979	0.0766
Tx12	16	0.29976	0.046908	0.37516	0.2123
Tx13	3	-1.0202	-0.069845	0.57944	0.96059
Tx14	10	0.95016	0.11811	0.47851	0.023759
Tx15	22	0.3704	0.067638	0.33187	0.13241
Tx16	8	0.65193	0.0726	0.48789	0.090988
Tx17	7	-1.369	-0.14273	0.57826	0.99089
Tx18	17	0.035158	0.0056665	0.37605	0.46277
Tx19	2	-0.37609	-0.021041	1.082	0.63586
Tx20	3	0.71301	0.048816	0.88384	0.21007

Figure 3: Gene-set analysis of brain volume measurement GWAS data using gene sets derived from a transcriptional hallmarks model of AD

icant changes in the loss value beyond this point.

A parallel analysis of the trends between the training and validation accuracy, as well as the training and validation loss, reveals a remarkable congruence. This similarity in trajectory is indicative of a well-generalized model; the model performs well not only on the data it was trained on but also on unseen data. This balance between training and validation performance helps to mitigate the risks of overfitting, where a model learns the training data too well and performs poorly on unseen data. We verified in the testing data, the testing loss is 0.1023 and the testing accuracy is 0.9648.

5.2.2. CNN TRANSFER LEARNING

The VGG16 customized model was integrated and trained to serve as our classification task. It culminated in a nearly perfect performance on the training data with a loss of approximately $1.9007\text{e-}06$ and an accuracy of 1.0000. The validation performance was strong as well, with a loss of 0.3925 and an accuracy of 0.9414.

The evolution of model performance across epochs is particularly noteworthy. Compared to the CNN model employed in the initial phase of our research, the transfer learning model

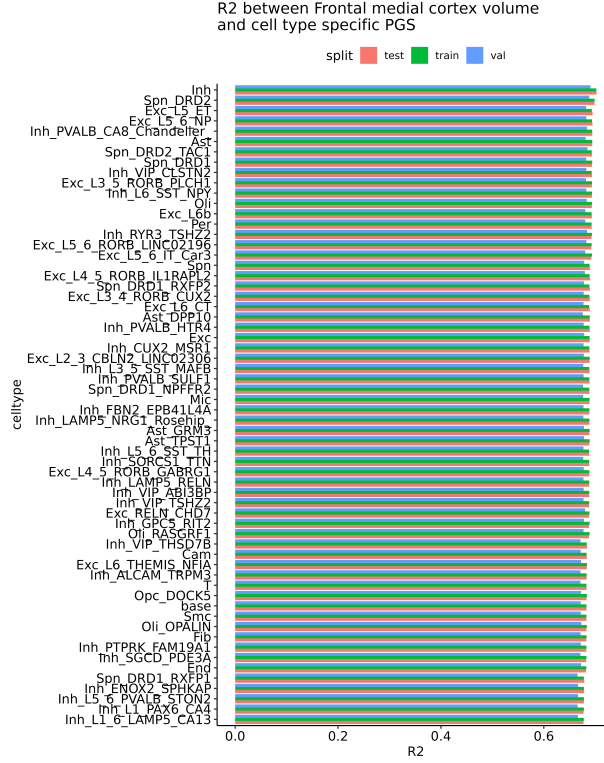


Figure 4: Inhibitory neurons, specifically the major cell type inhibitory peaks, provide a large scale enrichment for brain volume. Other cell types such as deep layer excitatory neurons and oligodendrocytes dominate the best-performing PGS.

using VGG16 showed a more pronounced change in the accuracy and loss metrics. This sharper ascent in accuracy and descent in loss led to the model achieving optimal performance as early as the 10th epoch. Subsequent to this point, the model’s accuracy for both training and validation data saw an increase and stabilized around the 15th epoch. This stability suggests that the model had successfully learned the primary features within the dataset and had minimized its error rate.

However, it is important to note the numerical difference in the accuracy and loss between the training and validation data. This discrepancy indicates a minor degree of overfitting, where the model exhibits superior performance on the training data compared to the validation data. Despite this, the overall performance of the model is robust and it provides a dependable tool for AD prediction.

5.2.3. VISION TRANSFORMER

The training process for our ViT model was halted at the 254th epoch, triggered by the lack of significant improvement in validation accuracy and the early stopping mechanism that we set up. Our ViT model yielded impressive results, obtaining a test accuracy of

approximately 0.9857, which marks our best performance thus far when relying solely on imaging data. Notably, the progression of loss reduction and accuracy increase throughout the training process was particularly stable. This stability is indicative of the model’s capacity to learn effectively from the MRI image patterns.

Unlike our CNN-based model, our ViT model exhibited negligible signs of overfitting. This observation is inferred from the minimal discrepancy between the training and test accuracy, which suggests that the model’s predictive ability is consistently high across both seen and unseen data.

Model	Train Accuracy	Test Accuracy
CNN	0.992	0.965
Transfer Learning	1.00	0.931
ViT	0.988	0.986

Table 1: Brain MRI Image Analysis Results

5.3. Multimodal Analysis

5.3.1. GENOTYPE-CLINICAL-DEMOGRAPHIC MODALITIES

We investigated the prediction of AD patients using the phenotype dataset from the UK Biobank. To simulate two distinct data modalities, we regarded participants’ genotype principal components as a separate modality from other physical traits and biomarkers. We found that most models performed similarly after fine-tuning and accounting for the unbalanced dataset Figure 6, and that the inclusion of the genotype principal components does not significantly improve model performance. This is expected as the general summarization of participants’ phenotypes to 40 dimensions likely obscured genotype signals that might be associated with AD. In addition, pre-training a joint encoding using the ContIG framework offered no improvements to model performance as well Table 2.

We then examined the feature weights of the elastic net logistic regression model Figure 7, which showed that age has a dominant strong effect on AD prediction. Some lab biomarkers such as platelet counts have a significant association with AD prediction, concurring with previous studies that platelet counts is a diagnostic biomarker for AD (Gowert et al. (2014)). However, the effect size of all these features are much weaker. This reflects the real-world difficulty in diagnosing AD using lab biomarkers, such that AD is typically confirmed via brain MRI images.

5.3.2. SIMULATED IMAGE-GENE MODALITIES

We find that contrastive pre-training for image-gene multimodal data can reach a better accuracy and loss Figure 8. Interestingly, we observe slower convergence for the pre-trained model, however, it converges at a better end-point. When we investigate the principal components of the learned self-supervised feature representations, we find that our image



Figure 5: Model performance history plots for: CNN built from scratch(top), CNN with transfer learning(middle), Vision Transformer(bottom)

Model	Overall ACC	AUROC	1-FPR	TPR
Majority Class Baseline	0.948	0.500	0.000	1.000
Logistic Regression (lab)	0.706	0.832	0.700	0.808
Logistic Regression (lab+genotype)	0.708	0.832	0.702	0.799
Balanced RF (lab)	0.682	0.822	0.674	0.830
Balanced RF (lab+genotype)	0.680	0.820	0.671	0.826
MLP (lab)	0.680	0.820	0.672	0.826
MLP (lab+genotype)	0.708	0.821	0.705	0.769
ContIG	0.709	0.813	0.705	0.768

Table 2: A comparison of various models with the ContIG model showed very similar model performances across models.

embeddings are clustered by image labels, suggesting influence by the simulated genetic features due to the cross-modal loss (images are contrasted against genetic information). Our final test accuracy for the pre-trained classifier was 93.056% and 88.657% for the

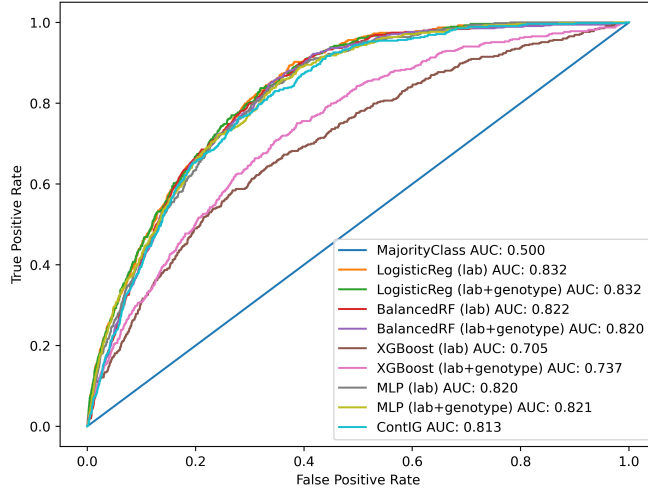


Figure 6: Summary of model performances on the prediction of Alzheimer’s disease using patient demographics, lab values and principle components of patient genotypes

baseline classifier. There was a $> 80\%$ true positive rate for every class, including minority classes and the pre-training improved glioma and meningioma tumor prediction greatly, as shown by Table 3. This suggests that the self-supervised contrastive pipeline would be a potentially effective strategy to predicting AD subtypes in the ROSMAP data given pre-training on the large UK Biobank data. Next, we investigate the SNP feature attributions.

Model	Overall ACC	$y_i = 0$ TPR	$y_i = 1$ TPR	$y_i = 2$ TPR	$y_i = 3$ TPR
ResNet Base	0.887	0.870	0.833	0.878	0.967
ContIG ResNet	0.931	0.916	0.940	0.857	0.967

Table 3: A comparison of baseline ResNet-18 performance and our self-supervised contrastive pre-trained ResNet-18 on simulated image-genetic test set. Accuracy and per-class true positive rates evaluated. All classes relatively equal size except $y_i = 2$, which is $1/3$ the training size.

from Integrated Gradients, on the image embeddings of the fine-tuned contrastively pre-trained classifier. From Figure 9, we observe that the for each label $y_i \in \{0, 1, 2, 3\}$, the most common set of 50 subsequent SNPs ($rs0$ to $rs49$, $rs50$ to $rs99$, etc.) is the set that was simulated to be more likely to be 2 for a given label (which was simulated in subsequent groups of 50). That is, for label 0, the most common set of subsequent SNPs is in sim_rs0 to sim_rs49 , for label 1, the most common set of subsequent SNPs is in sim_rs50 to sim_rs99 , and so on. This suggests that the final predictions may be somewhat informed by these simulated genetic data.

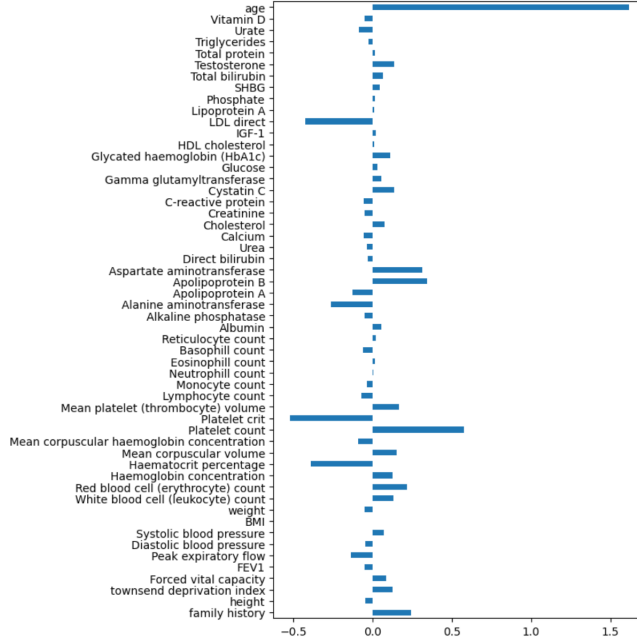


Figure 7: Coefficients of the logistic regression model shows that age has the strongest effect on AD.

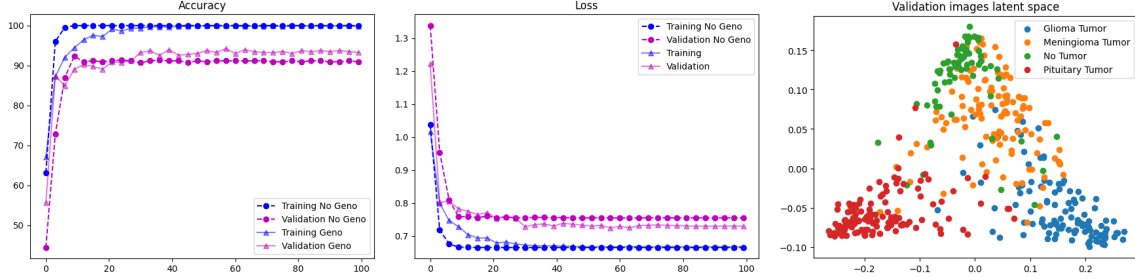


Figure 8: Training and validation accuracy and loss of multimodal self-supervised contrastive pre-trained ResNet18 (geno) and with no pre-training (no-geno) over 100 epochs (x -axis). The first two principal components of validation images in the joint embedding space is visualized (PC1 on x -axis, PC2 on y -axis).

6. Discussion

6.1. Genetics discussion topics

We were unable to obtain AD subtype genetics, so we mainly compared the feasibility of cell type specific PGS, and separately looking into the genetics of the subtypes via MAGMA in a gene-based fashion. Similarly, we only looked at one trait in the UK Biobank and found small but meaningful R^2 improvement in an ensemble PGS format. We imagine

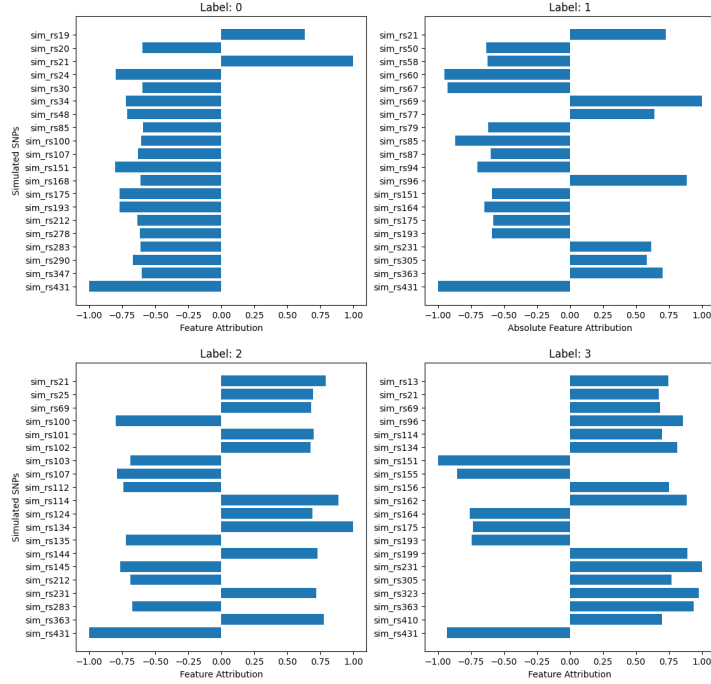


Figure 9: Top 20 mean absolute feature attributions of simulated genetic data on corresponding individuals’ MRI images within each class label.

that a scale-up of these methods and more rigorous analysis in the future over multiple sub-phenotypes may lead to better prediction with more mechanistic insights. We also did not compare to a raw thresholded model, where SNPs not in the peak set were discarded. We were advised against this because of a loss of genetic information rather than a hinting of better modeling. Additionally, due to computational constraints, many of the initial EpiMap PBMCs were not run to completion, so the first 50 iterations were used instead for comparison. But, we found in the end the neuronal PGS were the best performing.

6.2. Vision transformer outperforms CNNs for brain MRI imaging

We mainly compared the performance of vision transformers and CNNs. CNNs, especially transfer learning, suffer from overfitting problems. The validation and test accuracy is lower than training accuracy when using CNN-based architecture. For the plot of the vision transformer, the validation accuracy follows training accuracy closely in almost every epoch. The final test accuracy is almost the same as the training accuracy. This result implies that the vision transformer is capturing global and generalizable features which can prevent overfitting and be used in other datasets and tasks. As argued by [Jiang et al. \(2021\)](#), transformers can get rid of many ad-hoc building blocks commonly seen in CNN-based pipelines. Therefore, we believe that given enough computational resources and training data, we should be able to train a universal transformer model for brain MRI images like CheXNet for Chest

X-ray images developed by [Rajpurkar et al. \(2017\)](#).

The transformer model also has the potential to deal with 3D imaging data. Introduced by [Park and Kim \(2022\)](#), the idea of a vision transformer is to divide the image into patches and treat each patch like a ‘word’ in a ‘sentence’, which can be fed into a transformer. We can take this idea to divide a 3D $m * n * p$ image into several patches, flatten each patch to be a vector, similar to a word vector embedding representation, and feed the vector representations into the vision transformer. Limited by data availability, we weren’t able to implement this idea on 3D imaging data, which we originally planned to.

6.3. Cross-modal contrastive pre-training may improve AD subtype classification

These data, and the original claims in the ContIG paper [Taleb et al. \(2022\)](#), suggest that the cross-modal contrastive learning pipeline may be an effective pre-training method for image-GWAS data in the UK Biobank data when it is made available. The model could then be fine-tuned on our original task of AD subtype classification. Afterwards, we could investigate the genetic feature contributions to the various images corresponding to various subtypes. In addition, investigating reverse feature attributions may reveal which areas of the images are activated by given genetic data. Together, these suggest more interpretable and effective predictions that may improve the confidence in a clinician adopting this tool for patient diagnosis in practice. Another avenue of exploration is to use supervised contrastive learning, [Khosla et al. \(2021\)](#), instead of self-supervised in order to incorporate label information within the UK Biobank (like demographics, disease status, etc.). This may help further inform the pre-trained latent representations and result in better embeddings for subtype classification.

However, the image-gene contrastive learning analysis is limited by the simulated data. In reality, GWAS data doesn’t have neat correlations between outcomes and SNPs as the simulation implies. As such, we have no guarantees that the pre-training will result in downstream performance improvement. Furthermore, the original paper [Taleb et al. \(2022\)](#) investigates greater than two modalities and benchmarks against more models, so our performance may be hindered due to using only two modalities and one benchmark. Another limitation of this analysis is that it was done on proxy data due to access issues with UK Biobank. Hence, transferability may be an issue. Still, our pipeline shows promise for future AD subtype prediction and genetic correlations with MRI neuro-imaging.

7. Member Contributions

- Tony Ding - Brain MRI image data pre-processing and pipelining; CNN transfer learning and Vision Transformer(ViT) on brain MRI imaging data
- Benjamin James - Cell type-specific PGS
- Guangze Luo - Brain MRI image data pre-processing; CNN transfer learning for imaging data; pipeline for tabular data
- Ananth Shyamal - Gene-set analysis of brain volume measurement GWAS data using MAGMA
- Mathilde Tans - Brain MRI imaging data pre-processing
- Ryan Wang - Self-supervised contrastive pre-training pipelines/evaluation
- Boshen Yan - Data extraction and baseline AD model
- Chen Yang - Brain MRI imaging data pre-processing; CNN on brain MRI imaging data

8. Acknowledgements

We would like to express our deepest gratitude to the professor, Dr. Manolis Kellis, and our mentor, Yosuke Tanigawa for their guidance in the dataset, insightful critiques of methods, and encouragement throughout the course of this research. We also extend our appreciation to this course, which provided us with advanced knowledge with the opportunity to explore this fascinating area of study. Our thanks are also due to our excellent team members working together to contribute to the successful completion of this project.

References

- Heta Acharya, Rutvik Mehta, and Dheeraj Kumar Singh. Alzheimer disease classification using transfer learning. In *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, pages 1503–1508. IEEE, 2021.
- ADNI. Alzheimer’s disease neuroimaging initiative (adni), 2004, 2011. URL <https://adni.loni.usc.edu/>.
- Sartaj Bhuvaji. Brain tumor classification (mri), 2020. URL <https://www.kaggle.com/dsv/1183165>.
- Carles A Boix, Benjamin T James, Yongjin P Park, Wouter Meuleman, and Manolis Kellis. Regulatory genomic circuitry of human disease loci by integrative epigenomics. *Nature*, 590(7845):300–307, 2021.
- Jacob Boysen. Kaggle: Mri and alzheimers, 2017. URL <https://www.kaggle.com/datasets/jboysen/mri-and-alzheimers>.

- Chahd Chabib, Leontios J Hadjileontiadis, and Aamna Al Shehhi. Deepcurvmri: Deep convolutional curvelet transform-based mri approach for early detection of alzheimer’s disease. *IEEE Access*, 2023.
- Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L Yuille, and Yuyin Zhou. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021.
- Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations, 2020.
- Tom Heskes Danielle Posthuma Christiaan A. de Leeuw, Joris M. Mooij. Magma: Generalized gene-set analysis of gwas data. *PLOS Computational Biology*, 2015.
- CORDIS. Cordis: The final epad dataset is now available on the alzheimer’s disease workbench, 2021. URL <https://cordis.europa.eu/article/id/429468-the-final-epad-dataset-is-now-available-on-the-alzheimer-s-disease-workbench>.
- Data.gov. Data.gov: Alzheimer’s disease and healthy aging data, 2023. URL <https://catalog.data.gov/dataset/alzheimers-disease-and-healthy-aging-data>.
- Nikhil J Dhinagar, Sophia I Thomopoulos, Emily Laltoo, and Paul M Thompson. Efficiently training vision transformers on structural mri scans for alzheimer’s disease detection. *arXiv preprint arXiv:2303.08216*, 2023.
- Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.
- Guilherme Folego, Marina Weiler, Raphael F Casseb, Ramon Pires, and Anderson Rocha. Alzheimer’s disease detection through whole-brain 3d-cnn mri. *Frontiers in Bioengineering and Biotechnology*, 8:534592, 2020.
- Nina S. Gowert, Lili Donner, Madhumita Chatterjee, Yvonne S. Eisele, Seyda T. Towhid, Patrick Münzer, Britta Walker, Isabella Ogorek, Oliver Borst, Maria Grandoch, Martin Schaller, Jens W. Fischer, Meinrad Gawaz, Sascha Weggen, Florian Lang, Mathias Jucker, and Margitta Elvers. Blood Platelets in the Progression of Alzheimer’s Disease. *PLoS ONE*, 9(2):e90523, February 2014. ISSN 1932-6203. doi: 10.1371/journal.pone.0090523. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3938776/>.
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.
- Ehsan Hosseini-Asl, Georgy Gimel’farb, and Ayman El-Baz. Alzheimer’s disease diagnostics by a deeply supervised adaptable 3d convolutional network. *arXiv preprint arXiv:1607.00556*, 2016.

- Yifan Jiang, Shiyu Chang, and Zhangyang Wang. Transgan: Two pure transformers can make one strong gan, and that can scale up. *Advances in Neural Information Processing Systems*, 34:14745–14758, 2021.
- Edward Jung, Anshul Kashyap, Brandon Hsu, Mason Moreland, Chanon Chantaduly, and Peter D. Chang. A semi-supervised contrastive learning approach to alzheimer’s disease diagnostics using convolutional autoencoders. *medRxiv*, 2022. doi: 10.1101/2022.12.27.22283984. URL <https://www.medrxiv.org/content/early/2022/12/30/2022.12.27.22283984>.
- Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning, 2021.
- David S. Knopman, Helene Amieva, Ronald C. Petersen, Gäel Chételat, David M. Holtzman, Bradley T. Hyman, Ralph A. Nixon, and David T. Jones. Alzheimer disease. *Nature Reviews Disease Primers*, 7(1):33, May 2021. ISSN 2056-676X. doi: 10.1038/s41572-021-00269-y. URL <https://doi.org/10.1038/s41572-021-00269-y>.
- Narine Kokhlikyan, Vivek Miglani, Miguel Martin, Edward Wang, Bilal Alsallakh, Jonathan Reynolds, Alexander Melnikov, Natalia Kliushkina, Carlos Araya, Siqi Yan, and Orion Reblitz-Richardson. Captum: A unified and generic model interpretability library for pytorch, 2020.
- Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.
- Sachin Kumar and Sourabh Shastri. Alzheimer mri preprocessed dataset, 2022. URL <https://www.kaggle.com/dsv/3364939>.
- Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.
- Ruilin Li, Christopher Chang, Johanne M Justesen, Yosuke Tanigawa, Junyang Qian, Trevor Hastie, Manuel A Rivas, and Robert Tibshirani. Fast lasso method for large-scale and ultrahigh-dimensional cox model with applications to uk biobank. *Biostatistics*, 23(2):522–540, 2022.
- Cory Y McLean, Dave Bristor, Michael Hiller, Shoa L Clarke, Bruce T Schaar, Craig B Lowe, Aaron M Wenger, and Gill Bejerano. Great improves functional interpretation of cis-regulatory regions. *Nature biotechnology*, 28(5):495–501, 2010.
- OASIS. Open access series of imaging studies (oasis), 2007, 2010. URL <http://www.oasis-brains.org/>.
- Namuk Park and Songkuk Kim. How do vision transformers work? *arXiv preprint arXiv:2202.06709*, 2022.
- Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and

- Ilya Sutskever. Learning transferable visual models from natural language supervision. *CoRR*, abs/2103.00020, 2021. URL <https://arxiv.org/abs/2103.00020>.
- Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, et al. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225*, 2017.
- Saman Sarraf and Ghassem Tofghi. Classification of alzheimer’s disease using fmri data and deep learning convolutional neural networks. *arXiv preprint arXiv:1603.08631*, 2016.
- Saman Sarraf, Arman Sarraf, Danielle D DeSouza, John AE Anderson, Milton Kabia, and Alzheimer’s Disease Neuroimaging Initiative. Ovitad: Optimized vision transformer to predict various stages of alzheimer’s disease using resting-state fmri and structural mri data. *Brain Sciences*, 13(2):260, 2023.
- Shagun Sharma, Kalpna Guleria, Sunita Tiwari, and Sushil Kumar. A deep learning based convolutional neural network model with vgg16 feature extractor for the detection of alzheimer disease using mri scans. *Measurement: Sensors*, 24:100506, 2022.
- Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- Mukund Sundararajan, Ankur Taly, and Qiqi Yan. Axiomatic attribution for deep networks, 2017.
- Aiham Taleb, Matthias Kirchler, Remo Monti, and Christoph Lippert. Contig: Self-supervised multimodal contrastive learning for medical imaging with genetics. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20908–20921, June 2022.
- Yosuke Tanigawa. Single-cell transcriptional hallmarks and individual subtyping for alzheimer’s disease across 427 subjects., 2023.
- Yosuke Tanigawa, Junyang Qian, Guhan Venkataraman, Johanne Marie Justesen, Ruilin Li, Robert Tibshirani, Trevor Hastie, and Manuel A Rivas. Significant sparse polygenic risk scores across 813 traits in uk biobank. *PLoS Genetics*, 18(3):e1010105, 2022.
- Janani Venugopalan, Li Tong, Hamid Reza Hassanzadeh, and May D. Wang. Multimodal deep learning models for early detection of alzheimer’s disease stage. *Scientific Reports*, 11(1):3254, Feb 2021. ISSN 2045-2322. doi: 10.1038/s41598-020-74399-w. URL <https://doi.org/10.1038/s41598-020-74399-w>.
- Jade Xiaoqing Wang, Yimei Li, Xintong Li, and Zhao-Hua Lu. Alzheimer’s disease classification through imaging genetic data with ignet. *Frontiers in Neuroscience*, 16, 2022. ISSN 1662-453X. doi: 10.3389/fnins.2022.846638. URL <https://www.frontiersin.org/articles/10.3389/fnins.2022.846638>.
- Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How transferable are features in deep neural networks? *Advances in neural information processing systems*, 27, 2014.

Yuhao Zhang, Hang Jiang, Yasuhide Miura, Christopher D. Manning, and Curtis P. Langlotz. Contrastive learning of medical visual representations from paired images and text. *CoRR*, abs/2010.00747, 2020. URL <https://arxiv.org/abs/2010.00747>.