Toward Interpretable Brain Age Prediction and AD Classification

Rachel Iritani, Varnika Umashankar, Deborah Fu, Kuangchieh Lin, Chunyuan Hsu
University of Michigan, Ann Arbor
{riritani, varnika, debfu, kuanggay, chyhsu}@umich.edu

***Abstract*—As deep learning models have gained popularity in recent years, the capabilities of image processing algorithms have vastly improved. One way that this has a significant impact on society is within the realm of medical image processing. Advanced processing algorithms can aid in the understanding of biological mechanisms and in the monitoring and diagnosis of disease. Here, we focus on a specific application of medical image processing: brain age prediction from brain MRI. Convolutional Neural Networks (CNNs) can use the neuroanatomical features present in brain MRI to infer the subject's age. They can also be used to diagnose neurodegenerative disorders such as Alzheimer's disease (AD). Despite this, the CNN architecture limits the interpretability of the model, preventing us from gaining a thorough understanding of what features of the input were relevant to the predictions. Transformer models utilize an attention mechanism that allows the model to focus more on the most relevant parts of the input data, and allow for better interpretation of the model's decisions. This project focuses on optimizing a pipeline to analyze brain MRI, predict brain age, and perform AD classification, with a specific focus on interpretability. To do so, we are leveraging a pre-trained vision foundation model, with a transformer core, and designing regression and classification heads using the embeddings generated by the model. By creating a brain age prediction and AD classification paradigm that highlights the relevant features of the MRI data, we can gain a better understanding of neurodegeneration in healthy and diseased states and improve clinical detection of conditions like AD.**

I. Introduction

Image processing allows for the computational transformation of an image to identify the most meaningful features. Extracting contextually relevant information from images is a widely researched technique that is critical for image integration and computer vision implementations [1].

These kinds of methods are particularly impactful in identifying components of medical images relevant to a particular biological process or diseased state. This improves our understanding of biological mechanisms and can aid in the detection of a vast number of different conditions, often before symptoms are clinically detectable. While utilizing Magnetic Resonance Imaging (MRI) for brain age prediction is a well characterized problem [2], [3], current models tend to lack interpretability and clinical translatability required for practical medical use. By establishing a pipeline to identify features relevant to the predictions that the models are making, we can transform these algorithms from models strictly used for prediction to clinically informative platforms that can be integrated into existing workflows. Additionally, improving prediction and interpretability can provide a better understanding of the anatomical progression of neurodegenerative diseases such as Alzheimer's Disease (AD) and Parkinson's Disease (PD). This will also highlight biomarkers relevant for clinical detection. For this project, we plan to achieve this by building off of existing pretrained models to optimize brain age prediction and Alzheimer's disease classification. Specifically, our model framework will be designed to improve interpretability over current models in order to optimize the utility and clinical relevance of our analysis platform.

As previously mentioned, brain age prediction algorithms have been developed to identify the biological age based on neuroanatomical features detected from brain scans. With the increase in popularity of deep learning, architectures like Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) have been shown to have superior predictive accuracy in comparison to shallower models, at the cost of diminished interpretability [4]. While these models can be useful in characterizing neurological disorders like AD, their uninterpretable nature limits the explainability required for them to be trusted and adopted in a clinical environment. With the invention of transformers, interpretability in deep learning models is more feasible [5], [6]. In specific, a transformer architecture can be applied to a brain age prediction algorithm to identify which components of an MRI image are the most useful for predicting brain age. Similarly, this could be applied to an AD classification task to highlight which areas of an MRI image are most associated with AD diagnosis, suggesting what kinds of brain abnormalities are associated with AD. For this project, we chose to focus on Alzheimer's disease because 1) neurodegeneration, a feature of the disease, is easily detectable from MRI, and 2) AD is a highly prevalent condition. In specific, AD is the most common form of dementia and affects over 7 million Americans [7]. Despite this, there is still a gap in the understanding of the progression of the disease and accurate detection mechanisms remain sparse. By improving brain age prediction algorithms and downstream classification of disease state, we can develop models that will increase trust and understanding from clinicians and potential identify features that are relevant for better detection, thereby optimizing treatment and improving disease outcomes.

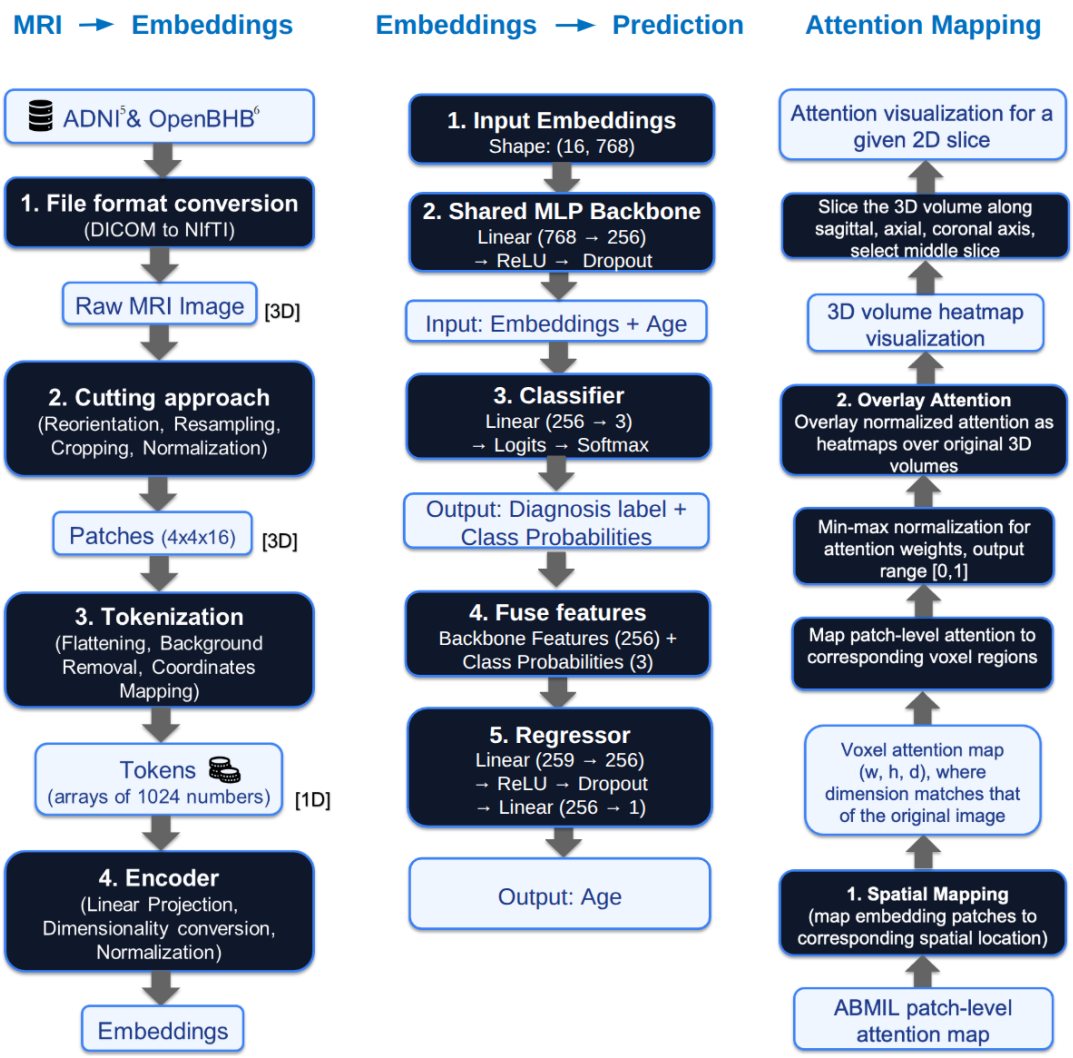

Fig. 1. Overall MRI preprocessing, embedding, prediction, and attention mapping pipeline.

II. Method

A. Dataset and/or Data Collection

We use brain MRI scans from the OpenBHB and ADNI datasets to support two downstream prediction tasks: Alzheimer's disease classification and brain age regression. These datasets provide structural MRI volumes together with subject-level demographic and clinical information.

For the OpenBHB dataset, the data collection process was relatively straightforward because the dataset provides direct download links for both MRI data and associated metadata. The metadata includes key subject-level variables such as age, sex, and imaging-related information, which can be directly matched with the downloaded MRI scans. Compared with ADNI, OpenBHB is easier to organize because its data access process is more centralized and its metadata structure is more consistent. In addition, since the OpenBHB scans come from a relatively smaller and more standardized set of sources, the preprocessing and normalization steps are more stable than in ADNI, where scans may come from many imaging sites with greater acquisition variability.

For the ADNI dataset, MRI scans were collected using the ADNI Analysis Ready Cohort Builder. We applied ARC filters to identify subjects with available structural brain MRI scans and corresponding clinical metadata. The selection focused on T1-weighted 3D MRI acquisitions, especially MPRAGE-related sequences, because these scans provide high-resolution anatomical information commonly used for Alzheimer's disease analysis. After filtering by imaging modality and acquisition description, we selected relevant MRI records and matched them with subject-level information such as subject ID, visit information, diagnosis, and age. When multiple MRI scans were available for the same subject, we prioritized standard 3D sagittal MPRAGE or accelerated MPRAGE acquisitions and selected one representative scan per subject to reduce duplicate sampling and maintain consistency across the cohort.

B. MRI to Embedding

To extract deep structural features from raw MRI scans, we use NeuroVFM, a 3D visual foundation model pre-trained on approximately 5 million clinical MRI and CT volumes. Instead of training a task-specific 3D CNN or transformer from scratch, we adopt a transfer learning approach and use NeuroVFM as a fixed embedding generator. This reduces computational cost and allows us to benefit from representations learned from large-scale medical imaging data.

Before embedding extraction, each MRI volume is processed through the NeuroVFM preprocessing pipeline. The input MRI can be provided as a NIfTI file, DICOM file, or study directory. Each image is loaded with SimpleITK, reoriented to a standard anatomical orientation, and spatially standardized. Specifically, the image is resampled to approximately 1 mm in-plane resolution and 4 mm through-plane resolution. The volume is then center-cropped so that its dimensions are divisible by the NeuroVFM patch size of $(4, 16, 16)$ voxels. After this spatial preprocessing step, the SimpleITK image is converted into a NumPy array with shape $[D, H, W]$. The MRI intensities are clipped using percentile-based thresholds and then min-max normalized to the range $[0, 1]$. These preprocessing steps help reduce variation caused by scanner differences, image orientation, voxel spacing, and intensity scale.

After preprocessing, each MRI volume is divided into 3D patch tokens. Each patch has size $4 \times 16 \times 16$, so each flattened patch token has length 1024. Background patches are removed by default, meaning that only foreground anatomical patches are passed into the encoder. The resulting batch contains the patch tokens, their 3D coordinates, sequence length information, and image metadata. This is important because NeuroVFM does not directly output one single vector for the whole MRI scan. Instead, it outputs a sequence of token-level embeddings, where each token corresponds to a local 3D patch of the MRI volume.

Inside the NeuroVFM encoder, each 1024-dimensional patch token is first standardized using MRI-specific mean and standard deviation values. The token is then projected into a transformer feature space and combined with 3D positional information based on its patch coordinate. The ViT backbone performs self-attention over the patch sequence and outputs a high-dimensional embedding for each foreground MRI patch. In the typical NeuroVFM configuration, the final embedding dimension is 768, so the output for one MRI scan has shape approximately:

$$\text{number\_of\_foreground\_tokens} \times 768. \tag{1}$$

For example, one scan may produce an embedding matrix such as 816×768, rather than a single 768-dimensional vector.

C. Embeddings to Prediction

Because the encoder produces token-level embeddings, a downstream aggregation step is required before scan-level prediction. We use a Attention-Based Multiple Instance Learning (ABMIL) framework adopted from the NeuroVFM pipeline. Each patch-level embedding acts as an instance, and the full set of patches from one image forms a bag of instances. Since the number of patch-level embeddings varies across MRI images, ABMIL helps in aggregating them into a fixed-size representation.

ABMIL computes an attention score for each patch embedding. These scores are calculated before pooling and represent how important each patch is for prediction. They are normalized using softmax within each image to produce attention weights, and the final image-level embedding is computed as a weighted sum of all patch-level embeddings.

The pooled patch-level embeddings (of dimension $D = 768$) are passed into a Multilayer Perceptron with different prediction heads. For Alzheimer's Disease classification, the MLP model outputs logits for diagnostic classes. For Brain Age regression, the MLP model outputs predicted brain age. In a joint-task setting, both classification and regression outputs are produced from the same pooled representation, and the model is trained on ground-truth disease labels for classification, and true brain ages for regression. The MLP itself consists of a shallow fully-connected architecture with one layer of size 128 or 256, followed by a non-linear activation (ReLU), dropout 0.1, and a final linear layer producing task-specific outputs.

To study how these tasks should interact with each other, we test three MLP variants. In MLPv0, classification and regression is performed using separate heads. In MLPv1, the classification probabilities are used to weight the regression outputs, which enforces a dependency on the predicted class. In MLPv2, the classification probabilities are concatenated with the pooled representations and used as an additional input to the regression head. This conditioning uses predicted class probabilities rather than ground-truth disease labels, allowing the model to learn from the classification signals while both tasks remain supervised independently. This forms a sequential training setup trained on joint loss objective with both Mean Squared Error loss for regression, and Cross Entropy Loss for classification. In this sequential training setup, the classification informs regression to make better predictions.

Overall, this pipeline converts raw structural MRI volumes into compact and expressive NeuroVFM-derived representations. These representations preserve local 3D anatomical information through patch-level tokens while also benefiting from large-scale pretraining. This makes the approach suitable for both Alzheimer's disease classification and brain age prediction.

D. Hyperparameter Tuning and Other Experiments

We performed hyperparameter tuning and evaluated different attention scoring functions. We searched over hidden_dim $\in \{128, 256\}$, learning_rate $\in \{10^{-3}, 10^{-4}\}$, batch_size $\in \{8, 16\}$, along with weight_decay $\in \{0.0, 10^{-4}\}$ for single-task models, and classification_loss_weight $\in \{1.0, 2.0\}$ for joint-task models. We also compared multiple attention scoring functions (gated, non-linear (ungated), linear, 2-layer MLP-based, and temperature-scaled gated), which differ in how

patch-level importance is computed before the softmax weighting step.

To mitigate overfitting, we used an early stopping mechanism which looked at 10-20 epoch windows, dropout of 0.1, class imbalance weighting in the classification loss, and tuned the classification loss' weight (classification_loss_weight).

E. Attention-Interpretability Analysis

NeuroVFM-generated embeddings (num_patches, 768) are aligned with (num_patches, 3) MRI volume spatial coordinates. Patch-level attention weights are then mapped onto a 3D voxel grid using the coordinates. The resulting map is a 3D voxel attention map of dimensions (width, height, depth), matching the dimensions of the original images. Note that since the attention values were calculated on a patch-level, the resolution of the voxel attention map reflects patch-level resolution. Optional interpolation may be performed, but we elected to simply expand the patch-level attention weights to overlay the corresponding $(4, 16, 16)$ voxel regions. Min-max normalization is performed for each volume to ensure attention values in the range of $[0, 1]$. The normalized attention map is overlaid as a heatmap over the original 3D MRI volume. For interpretability, the volumes are sliced along the sagittal, axial, and coronal axes, and the middle slices are selected for visualization of model decisions.

III. Related Work

Existing neuroimaging methods for brain MRI can be broadly grouped into three lines of work: task-specific model architecture design, data and benchmarking efforts, and pretrained or foundation-model approaches [2], [3]. This framing is consistent with our earlier proposal, which emphasized that traditional neuroimaging pipelines usually train separate supervised models for each downstream task rather than learning a single reusable representation.

First, architecture-focused studies mainly investigate how neural networks can capture age-related or disease-related structural information from MRI. In this setting, CNN-based approaches remain a strong baseline because they are effective at modeling local anatomical patterns in volumetric scans, and prior work has shown that such models can be useful for brain age estimation and for studying associations between brain age gaps and neurodegenerative disease [3], [6], [8]. More recent work has also incorporated transformers or hybrid CNN-transformer designs to better model long-range dependencies and global brain context, with the goal of improving robustness across heterogeneous MRI data settings [2], [7].

Second, beyond model design, an important line of existing work focuses on datasets, benchmarks, and generalization. Large multi-site neuroimaging datasets and comparative evaluations have shown that brain age prediction performance can vary substantially across data sources, preprocessing pipelines, and acquisition environments [9]. These findings suggest that models trained in a narrow or single-dataset setting may not generalize reliably, which makes transferable representations especially valuable in neuroimaging. Our previous proposal highlighted this issue as a key motivation for using broader pretrained representations rather than relying only on a task-specific supervised model.

Third, recent work has increasingly moved toward pretrained and foundation-model approaches for medical imaging and neuroimaging [8], [9]. These methods aim to learn general-purpose feature representations from large-scale imaging or multimodal corpora and then adapt them to downstream tasks. This direction is particularly attractive for brain MRI because annotation is expensive, 3D volumes are high-dimensional, and many clinically relevant tasks share common neuroanatomical structure. In our earlier proposal, this was the main motivation for using a pretrained 3D vision transformer, possibly NeuroVFM, as the base model rather than developing a completely new model from scratch.

Our current approach is similar to this line of work because we also rely on pretrained MRI representations instead of training an end-to-end task-specific architecture from scratch. More specifically, based on our current project direction, we use NeuroVFM as an embedding generator and then train downstream prediction models on top of those embeddings. In that sense, our work is most closely related to transfer-learning and foundation-model adaptation approaches in neuroimaging. This is also consistent with the direction stated in the proposal, which described leveraging a pretrained 3D MRI model as the starting point for downstream prediction.

At the same time, our project is narrower and more targeted than many existing foundation-model papers. We are not proposing a new large-scale foundation model, a new multimodal pretraining strategy, or a full diagnostic foundation pipeline. Instead, our focus is on evaluating whether pretrained NeuroVFM embeddings are useful for two specific downstream tasks: distinguishing Alzheimer's disease from normal brain MRI and predicting brain age. The brain age task is especially meaningful because the difference between predicted age and chronological age can potentially reflect abnormal aging patterns, which is relevant to neurodegenerative disease analysis. This targeted setup differs from broader prior work that aims to build fully general neuroimaging foundation systems or multimodal image-text alignment frameworks [10], [11].

A further difference is that our contribution is centered on downstream use of pretrained embeddings, not on introducing a new base architecture, and by explicitly emphasizing interpretability for clinical relevance alongside predictive performance.

IV. Experimental Results

A. Hyperparameter Tuning and Other Experiments

Hyperparameter tuning established strong single-task baselines on a held-out test set of 20% of the overall training data. For classification, the best configuration was hidden_dim = 256, batch_size = 8, learning_rate = 3×10^{-3}, and achieved a classification accuracy of 0.881 and macro F1 of 0.803. For regression, the best configuration was hidden_dim = 128,

batch_size = 8, learning_rate = 10^{-3}, and achieved a MAE of $\sim$3.95 and R2 of 0.959.

In the joint-task setting, the performance depended on how the tasks are connected. In MLPv0, the results were balanced but suboptimal with an accuracy of 0.884, F1 macro of 0.759, and MAE of 4.10, R2 of 0.953. MLPv1 improves regression to MAE of 3.79, R2 of 0.958, but worsens classification with F1 macro 0.57, accuracy of 0.749. MLPv2 achieves the best tradeoff and maintains classification accuracy of 0.873, and macro F1 0.775, while improving regression MAE to 3.54, and R2 of 0.966. (Table I)

We also evaluated the different attention scoring functions under the best-performing dual task model, MLPv2. Gated attention performed the best for classification, while the ungated non-linear attention performed the best under regression. (Table II)

B. Interpretability Analysing Using Attention Mapping

In order to better understand how the model was making its decisions, we extracted the attention maps generated from the different MLPs (Table II). First, we examined the differences between attention maps generated from the various attention mechanisms that were tested (Figure 2). This analysis demonstrated that the attention function with the gating mechanism removed (non-linear) learns the most expressive attention maps across the brain, as shown by the varying intensities of the color map displayed in Figure 2B. This is true across cognitively normal (CN), mild cognitively impaired (MCI), and Alzheimer's (AD) subjects. The non-linear model performed the best at the brain age prediction task, suggesting that attention heterogeneity is useful for the regression task. The temperature scaled gating mechanism appeared to have the least varying attention distributions, as shown by the very few brighter colored patches in Figure 2C. Generally, the dual MLP tasks tend to focus on regions near the prefrontal cortex and the hippocampus, aligning with previously established research that these regions are highly affected by AD and often the first to experience atrophy [12]. Additionally, some of the learned attention maps are asymmetrical, implying that the model has weighted a single hemisphere over another in its decision-making process, also aligned with the notion that AD often exhibits asymmetrical pathology [13], [14]. This suggests that this dual-prediction task and attention mapping back to the original input has the potential to highlight regions of high potential that could be of interest to clinicians and inform intervention.

Next, we compared the attention maps generated from the dual task, the regression task only, and the classification task only. This helped to provide a better understanding of the differences between the tasks and highlight potential reasons that the dual task is performing better overall. As can be seen in Figure 3, the dual task appears to generate attention distributions where high attention is more localized, whereas the single regression and classification tasks assign fairly constant weighting across the entire brain. This is likely because the dual task is trained sequentially, so the attention map is updated as a result of learning from both tasks, generating weights that are more informative and focus on the critical features of the images. This also shows that the sequential training paradigm is better for learning specific features, in contexts where interpretability is a priority.

Lastly, we examined the averaged attention maps across AD conditions that were generated from the dual task to better understand how consistent the learned attention is across subjects in a single group. From the averaged attention maps (Figure 4), it appears that the weighting between groups is fairly consistent, as distinct regions have significantly more weight magnitude than others. In all of the groups, the hippocampus region appears to have the highest attention, with both hemispheres having somewhat consistent weighting. In the CN and dementia groups, other regions in the brain including the frontal and parietal lobes seem to be of high relevance to the model's decision making. In the MCI group, only the hippocampal region seems particularly relevant, potentially because the hippocampus is often the only region to show pathology in MCI [12].

TABLE I
COMPARISON OF BEST PERFORMING HYPERPARAMETERS OF DIFFERENT MLP ARCHITECTURES IN JOINT-TASK

Model	hidden_dim	batch_size	learning_rate	F1 macro	Accuracy	MAE	R2
MLPv0	256	16	10^{-3}	0.759	0.884	4.10	0.953
MLPv1	128	8	10^{-3}	0.569	0.749	3.79	0.958
MLPv2	256	16	10^{-3}	0.775	0.873	3.54	0.966

TABLE II
EFFECT OF ATTENTION SCORING FUNCTION ON MLPV2

Attention type	F1 macro	Accuracy	MAE	R2
Gated	0.779	0.876	3.56	0.965
Nonlinear	0.748	0.865	3.51	0.967
Linear	0.742	0.865	3.66	0.964
MLP	0.774	0.887	3.99	0.960
Temperature-scaled gated	0.763	0.873	3.69	0.963

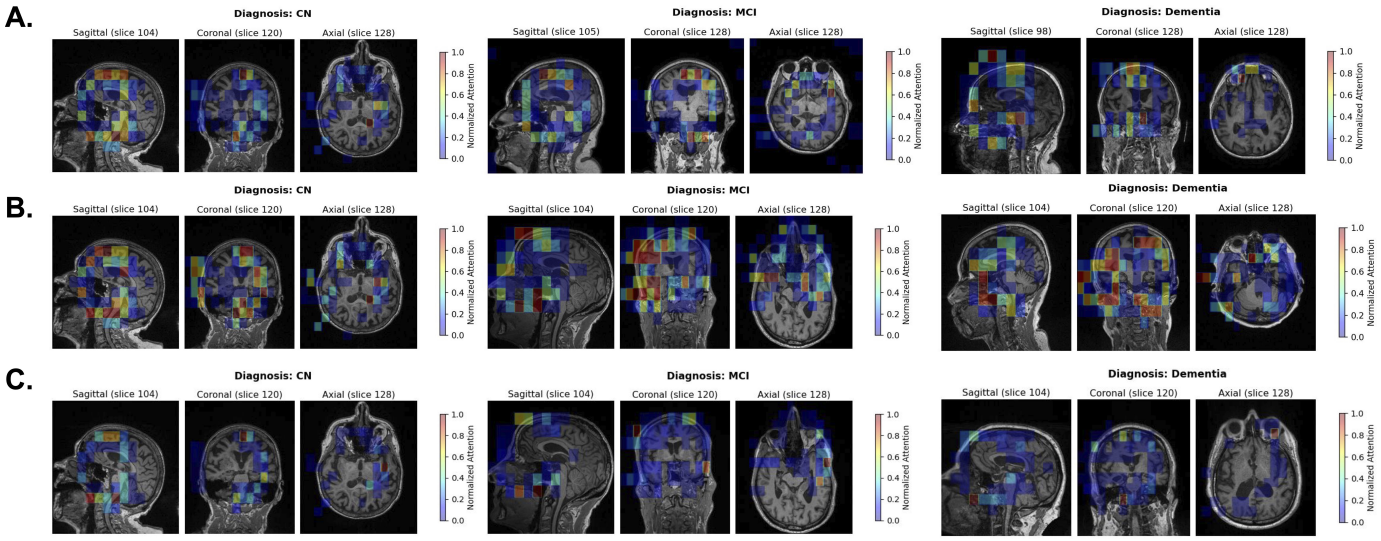

Fig. 2. Generated attention maps overlaid on MRI slices using A) Baseline attention mechanism (gating) B) Non-linear attention (no gating) and C) Temp-scaled gated attention.

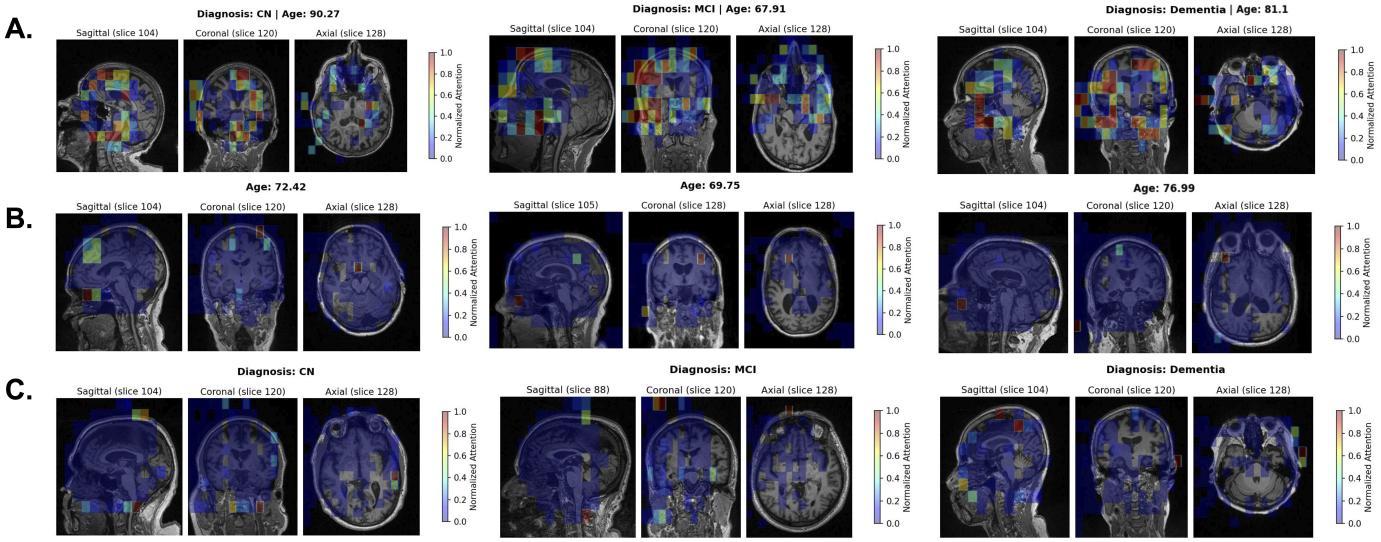

Fig. 3. Generated attention maps overlaid on MRI slices from A) Dual sequential classification-regression task, B) Regression task only, C) Classification task only.

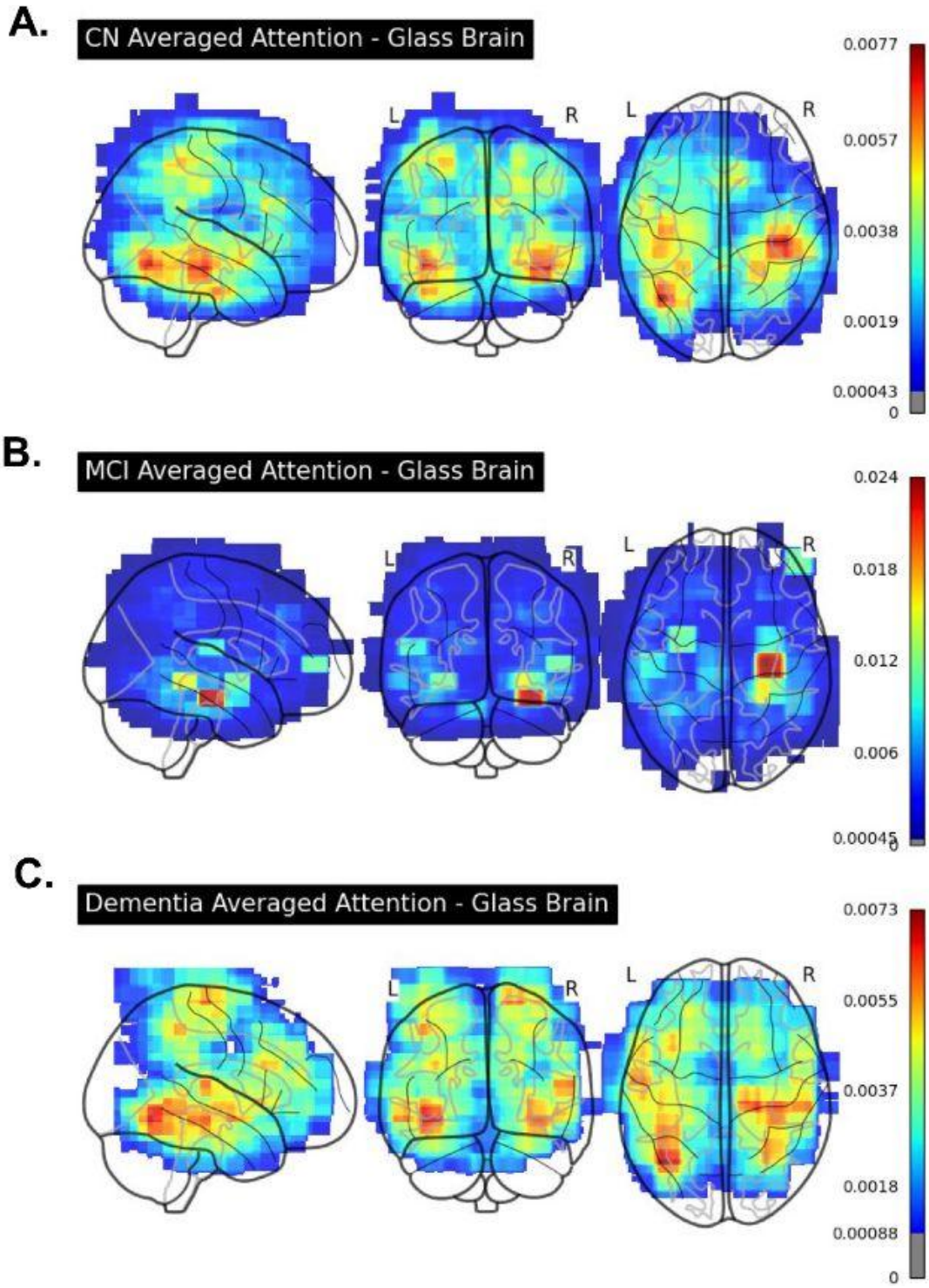

Fig. 4. Averaged attention maps across AD diagnosis conditions A) cognitively normal, B) mild cognitively impaired and C) dementia.

V. Discussion

The key finding from the hyperparameter tuning with different architectures is that how classification and regression tasks are connected affects their individual performances. In MLPv0, no shared information gives baseline results, whereas in MLPv1, the classification performance is actually affected negatively. MLPv2, gives a better balance, allowing the classification task's outputs to guide the regression task to make better predictions. This indicates that the classification task's outputs, i.e., the class probabilities carry information useful for regression analysis. Currently, the setup is unidirectional, from classification to regression, but this work can be extended to make the model architecture bidirectional such that the regression task also informs classification.

While the attention scoring mechanism does influence results, the effect is smaller than that of the model design choices. Overall, improving how the tasks share information is more important for improving performance than refining the aggregation step alone.

VI. Conclusion

In this project, we developed an interpretable MRI analysis pipeline for brain age prediction and Alzheimer's disease classification using pretrained NeuroVFM embeddings. Raw structural MRI scans from OpenBHB and ADNI were converted into standardized volumetric inputs, processed through the NeuroVFM preprocessing pipeline, and represented as patch-level transformer embeddings. Because NeuroVFM outputs a sequence of local patch embeddings rather than a single scan-level vector, we used an attention-based multiple instance learning framework to aggregate patch-level features into image-level representations for downstream prediction.

Our results show that NeuroVFM embeddings are effective for both tasks. Single-task models achieved strong classification and regression performance, while the joint-task MLPv2 architecture provided the best overall balance between AD classification and brain age prediction. In particular, MLPv2 improved regression performance to an MAE of 3.54 and R^2 of 0.966 while maintaining competitive classification accuracy. These results suggest that disease classification information can help inform brain age prediction when the tasks are connected in a flexible way.

The interpretability component of the project is a key contribution. By mapping patch-level attention weights back onto the original MRI volume, we generated attention heatmaps that visualize which regions of the brain contributed most to the model's predictions. Although these attention maps should not be interpreted as causal clinical explanations, they provide a useful first step toward understanding model behavior and identifying anatomically relevant regions.

In summary, this project demonstrates that pretrained 3D visual foundation models can be adapted for clinically relevant neuroimaging tasks with relatively lightweight downstream models. The combination of NeuroVFM embeddings, attention-based pooling, joint-task learning, and MRI-space attention visualization provides a promising framework for accurate and interpretable brain age prediction and AD classification. Future work should focus on larger datasets, stronger external validation, sharper attention localization, and deeper analysis of whether the predicted brain age gap can serve as a meaningful biomarker for neurodegeneration.

Author Contributions

VU: helped with experimental design, designing and training the MLP model and the attention scoring mechanism, performing hyperparameter tuning and other experiments, generating and analyzing experimental results and contributed in writing sections 2, 4, and 5.

Chyhsu: helped with data collection, embeddings and coordinations generating and contributed in sections 2(data collection, MRI to embeddings), 6 and organizing a final version of the final report

Rachel: wrote abstract + intro, helped with initial MLP design and training, did attention mapping and visualization, wrote attention results

REFERENCES

[1] S. Hallur and A. Gavade, "Image feature extraction techniques: A comprehensive review," *Franklin Open*, vol. 12, p. 100366, Sep. 2025.

[2] S. Malla and S. R. Dutta, "NeuroAgeFusionNet: An ensemble deep learning framework integrating CNN, transformers, and GNN for robust brain age estimation using MRI scans," *Scientific Reports*, vol. 15, 2025.

[3] H. Kim *et al.*, "A novel deep learning-based brain age prediction framework for routine clinical MRI scans," *npj Aging*, vol. 11, no. 1, p. 70, Jul. 2025.

[4] A. Lombardi, D. Diacono, N. Amoroso, A. Monaco, J. M. R. S. Tavares, R. Bellotti, and S. Tangaro, "Explainable deep learning for personalized age prediction with brain morphology," *Frontiers in Neuroscience*, vol. 15, p. 674055, 2021.

[5] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[6] B. Kotipalli, "The role of attention mechanisms in enhancing transparency and interpretability of neural network models in explainable AI," Harrisburg University, 2024. [Online]. Available: https://digitalcommons.harrisburgu.edu/dandt/2

[7] Alzheimer's Association, "Facts and figures," 2020. [Online]. Available: https://www.alz.org/alzheimers-dementia/facts-figures

[8] J. Kim, M. Kim, and H. Park, "Domain aware multi-task pre-training of 3D Swin transformer for brain MRI," in *Proceedings of the Asian Conference on Computer Vision*, 2024.

[9] R. P. Dörfel *et al.*, "Prediction of brain age using structural magnetic resonance imaging: A comparison of clinical utility of publicly available software packages," *eBioMedicine*, vol. 123, p. 106094, 2026.

[10] S. Zhang *et al.*, "A multimodal biomedical foundation model trained from fifteen million image–text pairs," *NEJM AI*, vol. 2, no. 1, 2024.

[11] A. Kondepudi *et al.*, "Health system learning achieves generalist neuroimaging models," arXiv:2511.18640, 2025.

[12] D. Mrdjen, E. J. Fox, S. A. Bukhari, K. S. Montine, S. C. Bendall, and T. J. Montine, "The basis of cellular and regional vulnerability in Alzheimer's disease," *Acta Neuropathologica*, vol. 138, no. 5, pp. 729–749, Nov. 2019.

[13] T. E. Anijärv, R. Ossenkoppele, R. Smith, A. Pichet Binette, L. E. Collij, H. H. Behjat, J. Rittmo, L. Karlsson, K. Ahmadi, O. Strandberg *et al.*, "Hemispheric asymmetry of tau pathology is related to asymmetric amyloid deposition in Alzheimer's disease," *Nature Communications*, vol. 16, no. 1, p. 8232, Sep. 2025.

[14] Y. Pusparani, C. Y. Lin, Y. K. Jan, F. Y. Lin, B. Y. Liau, J. S. R. Alex, J. Aparajeeta, W. H. Chao, and C. W. Lung, "Hippocampal volume asymmetry in Alzheimer disease: A systematic review and meta-analysis," *Medicine*, vol. 104, no. 10, p. e41662, Mar. 2025.