Multi-Modal Deep Feature Integration for Alzheimer's Disease Staging

Amritpal Singh
College of Computing
Georgia Institute of Technology
Atlanta, USA
asingh880@gatech.edu

Wenqi Shi
Department of Electrical Engineering
Georgia Institute of Technology
Atlanta, USA
wshi83@gatech.edu

May D. Wang
Department of Biomedical Engineering
Georgia Institute of Technology
Atlanta, USA
maywang@gatech.edu

***Abstract*—Alzheimer's disease (AD) is one of the leading causes of dementia and 7th leading cause of death in the United States. The provisional diagnosis of AD relies on comprehensive examinations, including medical history, neurological and psychiatric examinations, cognitive assessments, and neuroimaging studies. Integrating diverse sets of clinical data, including electronic health records (EHRs), medical imaging, and genomic data, enables a holistic view of AD staging analysis. In this study, we propose an end-to-end deep learning architecture to jointly learn from magnetic resonance imaging (MRI), positron emission tomography (PET), EHRs, and genomics data to classify patients into AD, mild cognitive disorders, and controls. We conduct extensive experiments to explore different feature-level and intermediate-level fusion methods. Our findings suggest intermediate multiplicative fusion achieves the best stage prediction performance on the external validation dataset. Compared with unimodal baselines, we can observe that integrative approaches that leverage all four modalities demonstrate superior performance to baselines reliant solely on one or two modalities. In an age-wise comparison, we observe a unique pattern that all fusion methods exhibited superior performance in the earlier age brackets (50-70 years), with performance diminishing as the age group advanced (70-90 years). The proposed integration framework has the potential to augment our understanding of disease diagnosis and progression by leveraging complementary information from multimodal patient data.**

***Index Terms*—Multimodal Integration, Stage Prediction, Feature Fusion, Joint Representation, Alzheimer's Disease**

I. Introduction

Alzheimer's disease (AD) represents a continually advancing neurological disorder that currently stands as the seventh leading cause of mortality in the United States [1]. Six million individuals in America alone are burdened by the detrimental impacts of Alzheimer's disease, with projections indicating a potential tripling of this figure by the middle of the century [1]. The projected annual financial implications related to patient care are staggering, approximating 305 billion [2]. Due to the lack of prevailing absence of definitive cures, it is important to facilitate early detection and intervention to mitigate the progression of the disease [3].

Clinicians commonly utilize both medical imaging and clinical data for the early detection of disease onset [3].

This research has been supported by a Wallace H. Coulter Distinguished Faculty Fellowship, a Petit Institute Faculty Fellowship, and research funding from Amazon and Microsoft Research to Professor May D. Wang.

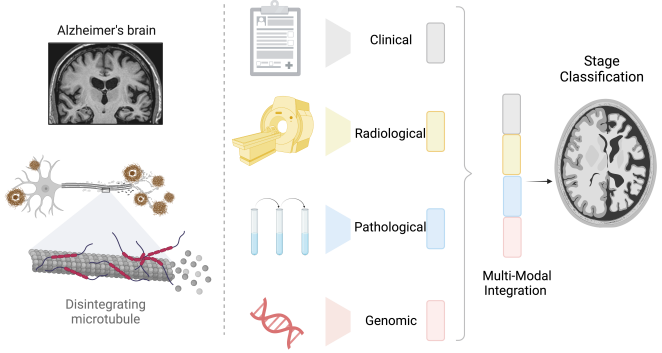


Fig. 1: Overview of multi-modal deep feature integration model for AD stage prediction. Created with BioRender.com.

Among imaging modalities, Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET) are prevalently employed modalities for AD detection and staging. MRI assists in the identification of cerebral atrophy associated with dementia. Particularly in AD, MRI features progressively intensifying cerebral atrophy, initiating in the medial temporal lobe, inclusive of the entorhinal cortex, hippocampus, amygdala, and parahippocampus. The progression of atrophy on MRI closely mirrors the histopathological spread of neurofibrillary tangles. While MRI can effectively identify these volume atrophy alterations, the atrophy patterns tend to overlap between various diseases, making it an unspecific biomarker for AD. Furthermore, MRI lacks molecular specificity and cannot detect the histopathological hallmarks of AD, such as amyloid plaques or neurofibrillary tangles. On the other hand, PET scans, specifically FDG-PET, use the glucose analog FDG to detect synaptic activity and brain metabolism. FDG-PET has displayed characteristic hypometabolism patterns primarily in limbic and association regions. Although FDG-PET can detect hypometabolism prior to the manifestation of cognitive symptoms, FDG retention is a non-specific indicator of metabolism and does not directly relate to AD pathology. Consequently, combining structural and metabolic biomarkers from MRI and PET scans can offer complementary information, thereby enhancing the detection of AD.

Recent advances in deep learning have resulted in an increased focus on AD staging. In addition, access to multi-

modal clinical data enables the deep learning models to leverage complementary information from multiple modalities simultaneously, enhancing the overall performance potential [6], [7], [8]. Bae et al. [7] employed 2D convolutional neural networks (CNNs) to execute binary classification tasks on 3D MRI data, supplementing this process with additional information regarding layer numbers and patient clinical attributes. Similarly, Venugopalan et al. [6] merged 3D CNN features from MRI with electronic health records (EHRs) and genomics SNP data, achieving a comparative accuracy of 87.2%. El-Sappagh et al. [8] integrated EHR data with manually extracted features from MRI and PET scans into a random forest model for AD stage classification. Meanwhile, Lu et al. [9] designed an end-to-end deep learning network specifically tailored for binary classification tasks, distinguishing Alzheimer's disease from MCI, utilizing MRI and PET scans as inputs. In addition, several existing studies leveraged manual feature extraction for disease classification, despite its potential to yield meaningful results, which can be labor-intensive and susceptible to distribution shifts over time [6]. As a result, utilizing an end-to-end training approach with automatic feature extraction plays an important role in facilitating more efficient model development for AD stage classification. Best to our knowledge, this is first comparative study of different multimodal fusion techniques to fuse medical imaging and EHR tabular data.

In this study, we propose different end-to-end multi-modal deep feature integration methods to predict disease stages for Alzheimer's disease and compare our results with previous state-of-the-art methods. We summarize our contributions in three fold:

- In this study, we propose different end-to-end multi-modal deep feature integration methods to combine MRI, PET, EHRs, and genomic data to jointly predict disease stages, including AD, mild cognitive impairment (MCI), and cognitive normal (CN).
- We investigate both feature-level and intermediate-level multi-modality fusion for deep learning architectures to jointly learn from multi-scale image and tabular data.
- In our experiments, we present both single-modality and multi-modalities model performance along with their 95% confidence intervals (CI) to demonstrate the effectiveness of the proposed integration models.

II. Materials and Methods

A. Data Description

The data utilized in this study is curated from the Alzheimer's Disease Neuroimaging Initiative (ADNI) [5] database (adni.loni.usc.edu). The foremost objective of ADNI has been to investigate the potential of combining MRI, and PET, among other biological markers, to enhance our understanding of the staging and progression of MCI and AD.

Scans from all clinical visits, ranging from initial screening to 48-month follow-ups, were incorporated. Upon exclusion of mask and segmentation scans from the dataset, we selected patients possessing at least one MRI and PET scan, along with their corresponding EHR and genomic data. Regarding the EHRs and genomics data, we extracted patient information from family history, physical examinations, neurological evaluations, and blood investigations.

TABLE I: Details of patient-level dataset split into the train, validation, and test sets.

Split	Patient	Modality	AD	CN	MCI	Total
Train	307	MRI	238	384	388	1010
		PET	437	626	518	1581
Val	39	MRI	33	29	58	120
		PET	60	48	79	187
Test	38	MRI	39	39	46	124
		PET	75	55	71	201

The final dataset contains multiple MRI and PET scans per patient, collected over multiple visits. On average, each patient contributed approximately 3.2 MRI and 5.12 PET scans. To prevent potential data leakage, a patient-level division was implemented with an 80:10:10 ratio allocated for the training, validation, and testing datasets, respectively. A comprehensive breakdown of the distribution can be found in Table I.

We present more feature details with demographics of EHRs and genomics data from ADNI in Table. II.

TABLE II: EHR and genomic feature details from source ADNI dataset.

Data	Feature details
FamilyHist.csv	FHQMOM, FHQMOMAD, FHQDAD, FHQDADAD
PatientSymp.csv	BCNAUSEA, BCVOMIT, BCDIARRH, BCCONSTP, BCABDOMN, BCSWEATN, BCDIZZY, BCENERGY, BCDROWSY, BCVISION
neuroexam.csv	NXVISUAL, NXAUDITO, NXTREMOR, NXCONSCI, NXNERVE, NXMOTOR, NXFINGER, NXHEEL, NXSENSOR, NXTENDON, NXPLANTA, NXGAIT
Desikanlab.csv	PHS, CIR
Loclab.csv	CTWHITE, CTRED, PROTEIN, GLUCOSE
Labdata.csv	HMT97, HMT56, HMT57, HMT59, HMT60, HMT61, HMT62, HMT96, HMT63, HMT64, HMT65, HMT49, HMT48, HMT50, HMT52, HMT53, HMT54, HMT58, HMT70, HMT94, HMT72, HMT80, HMT81, HMT83, HMT84, HMT85, HMT86, HMT95, HMT87, HMT88, HMT89, HMT75, HMT74, HMT76, HMT77, HMT78, HMT79, HMT82, HMT98, HMT99, HMT21, HMT66, HMT67, HMT68, HMT69, HMT20, HMT90, HMT91, HMT92, HMT93, UAT65, HMT55
Ptdemog.csv	PTGENDER, PTMARRY, PTEDUCAT, PTETHCAT, PTRACCAT
Apoeres.csv	APVOLUME, APGEN1, APGEN2, APRECEIVE, APAMBTEMP, APRESAMP, APUSABLE
Physical.csv	PXGENAPP, PXHEADEY, PXNECK, PXCHEST, PXHEART, PXABDOM, PXEXTREM, PXPERIPH, PXSKIN, PXMUSCUL, PXOTHER, PXABNORM
Others	Sex, Age

B. Data Pre-Processing

Postprocessed brain images were created after additional processing like skull stripping, masking background, and im-

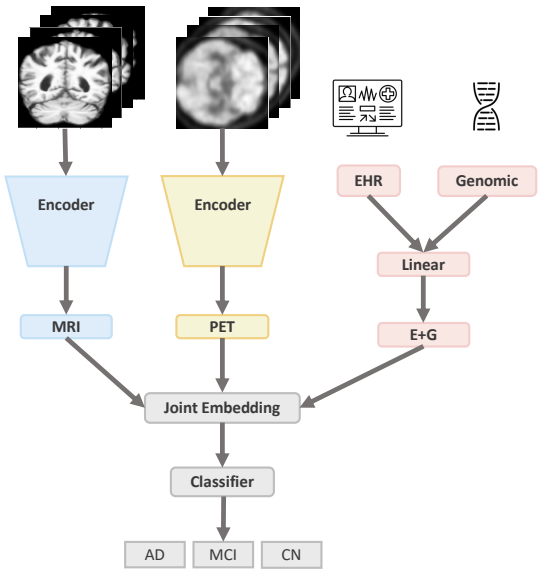


Fig. 2: Overview of intermediate-level fusion. First, we train an encoder to extract latent embeddings from MRI (blue), PET (yellow), EHRs, and genomics (red) for effective feature representation. We then apply intermediate-level fusion methods to generate the multi-modal joint embeddings for classification.

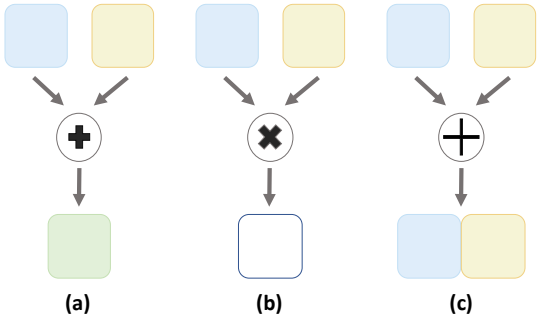


Fig. 3: Three different types of intermediate-level fusion methods. (a) Additive fusion, (b) multiplicative fusion, and (c) feature concatenation.

age registration. The default dimensions of the MRI images were set at $224 * 224 * 224$. We proceeded to eliminate slices devoid of brain regions along the x, y, and z axes and resized the image to a more manageable $224 * 224 * 16$ dimension. Subsequently, all voxel values within the image were normalized, culminating in an image with values ranging between 0 and 1.

For EHR and genomic data, standard scaling was employed for numerical variables, whereas categorical variables underwent one-hot encoding. This processing resulted in a final vector dimension of 318 for each individual patient.

C. Multimodal Deep Feature Integration

In this section, we present a multi-modal integration framework, to generate feature embeddings and deploy fusion for the integration of multimodal data. The architecture is fundamentally tripartite, consisting of feature extraction, fusion, and classifier training stages. Specifically, we explore two different feature integration approaches, intermediate-level and feature-level fusion, to develop a robust multimodal approach.

1) Unimodal Architecture: As baselines, we begin by training neural network models using individual modalities. For MRI and PET scans, we employ the 3D ResNet18 convolutional neural networks (CNNs) architecture to extract feature embeddings. In the case of EHR and genomics data, we leverage blocks of dense layers with RELU activation for latent embeddings. These embeddings are used to train unimodal classifiers for the classification of disease staging.

2) Intermediate-level Fusion: To leverage the combined information from MRI and PET scans, we use 3D Resnet18 CNNs to extract embeddings from each imaging modality. Embeddings from imaging modalities can be fused using different methods, such as concatenation, additive fusion, or multiplicative fusion. Fig. 2 shows the architecture of the intermediate-level fusion approach.

Formally, given data x = $(x^{(1)}, ...x^{(K)})$ containing K modalities, we can generate embeddings $e = (e^{(1)}, ...e^{(K)})$ with shape $(s^{(e^1)}, ...s^{(e^k)})$ for each modality, from their respective encoders f = $(f^{(1)}, ...f^{(K)})$. The goal of these encoders is to learn appropriate transformations and biases that capture the relationships between the modalities and the output variable. For concatenative fusion, we generate joint embeddings using X_{joint} = $(e^{(1)} \oplus ...e^{(K)})$, where $s^{X_{joint}} = s^{(e^1)} + ... + s^{(e^k)}$. For additive fusion, we generate joint embeddings with X_{joint} = $(e^{(1)} + ...e^{(K)})$, where $s^{X_{joint}} = s^{(e^k)}$. For multiplicative fusion, joint embeddings are extracted with X_{joint} = $(e^{(1)} \cdot ...e^{(K)})$, where $s^{X_{joint}} = s^{(e^k)}$. Fig. 3 illustrates a pictorial representation of the three fusion methods.

When combining MRI and PET modality, we utilize X_{joint} from the fusion of image embeddings directly to train the classifier. For the combination of MRI, PET, EHRs, and genomics modalities, we first concatenate joint embedding for imaging embeddings with EHRs and genomics embedding. Then, we leverage these combined embeddings for the classifier training.

3) Feature-level Fusion: In the context of feature-level integration, we first transform the EHRs and genomic data into an image-like representation. Figure 4 illustrates a visual depiction of the architecture employed for feature-level integration. Inspired by the DeepInsight approach [15], we employ PCA for feature selection, facilitating the conversion of normalized tabular data into a pixel representation. Subsequently, the resulting output is resized to align with the desired dimensions. Figure 5 showcases the resultant images generated from this process. The generated image-like representations, along with the MRI and PET images, are concatenated into a combined feature, giving rise to a unified 3D image. During the training phase, this composite image is fed into a 3D Resnet CNN to extract feature embeddings for the following classification.

4) Classification and Optimization: Lastly, we frame the AD staging as a multiclass classification and use a supervised learning schema to train the classification architecture. Given N training samples, with y_{ic} and $\hat{y}_{ic}$ indicating the ground truth and predicted outcome for i_{th} subject belonging to class

TABLE III: Main results on the hold-out test dataset with 95% confidence interval.

Modality	Fusion method	Acc	One vs One Acc			Specificity		
			AD vs CN	AD vs MCI	CN vs MCI	AD	CN	MCI
MRI	-	45.2±1.0	75.1±3.5	60.9±1.7	53.8±2.1	89.2±1.6	61.9±9.9	66.2±13.2
PET	-	57.5±1.1	80.0±5.5	72.1±2.0	63.5±2.0	91.4±4.7	77.8±3.1	66.1±8.2
EHRs+genomics	-	60.4±4.2	90.2±3.4	73.3±5.3	68.8±4.6	86.4±2.8	82.3±5.3	71.2±7.8
MRI+PET	Interm. Concat.	52.0±0.6	77.6±1.2	73.8±1.6	55.2±2.7	95.9±4.0	69.7±9.6	61.2±7.8
MRI+PET	Interm. Add.	54.6±4.8	77.1±1.8	74.0±4.7	59.0±3.6	95.9±0.02	78.8±0.04	59.0±3.6
MRI+PET	Interm. Multi.	55.1±3.6	77.4±2.7	73.1±4.4	60.7±4.8	96.4±1.8	77.6±8	57.1±13.3
MRI+PET+EHRs+genomics	Feature level	64.3±1.7	90.7±3.9	79.5±1.7	68.3±1.1	97.8±2.3	79.5±1.7	**68.3±1.2**
MRI+PET+EHRs+genomics	Interm. Concat.	62.6±2.2	87.7±5.1	**81.2±1.4**	63.6±3.6	96.6±1.2	80.9±16.0	65.1±16.9
MRI+PET+EHRs+genomics	Interm. Add	62.7±4.1	90.0±5.9	80.4±3.0	64.3±5.4	97.6±2.8	84.5±2.4	64.5±6.2
MRI+PET+EHRs+genomics	Interm. Multi.	**66.2±2.6**	**91.5±4.5**	79.2±2.8	**71.2±3.4**	**98.3±2.7**	**89.7±2.5**	59.7±2.1

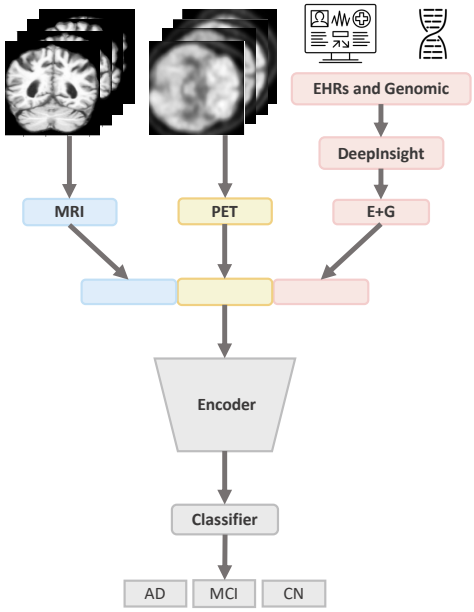


Fig. 4: Overview of feature-level fusion. EHR and genomics (red) are converted into pixel representation, combined with raw MRI (blue) and PET (yellow) scans to generate a single image volume.

Fig. 5: Examples of EHR and genomics data as represented images using DeepInsight [15].

c, we then calculate cross-entropy loss L_{ce} as:

$$L_{ce} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{ic}.$$

We use backpropagation with ADAM optimizer for model training and optimization. In addition, we use early stopping to prevent overfitting. For the model comparison and performance evaluation, we use overall accuracy, one-vs-one accuracy, and specificity for each class as evaluation metrics.

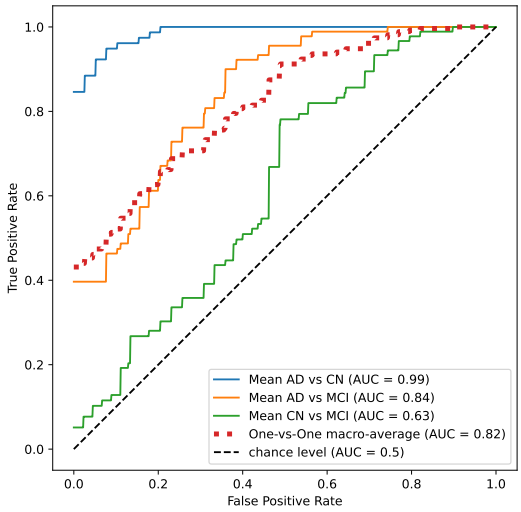


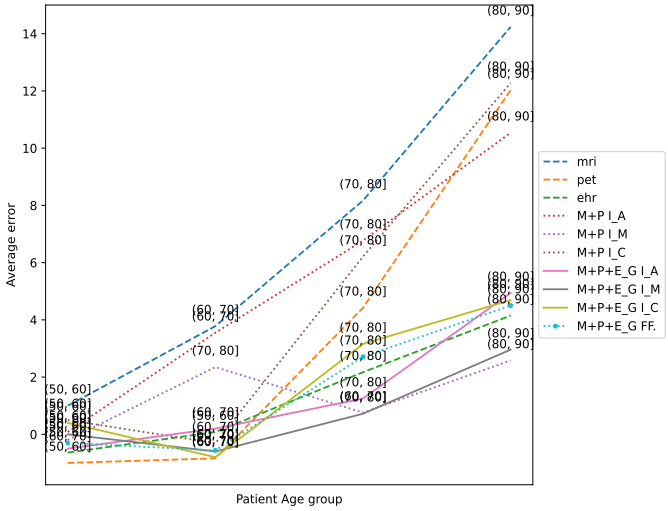
Fig. 6: ROC for the multiplicative fusion method on the hold-out test set.

We report the 95% confidence interval for each metric within 5 times of training for model robustness.

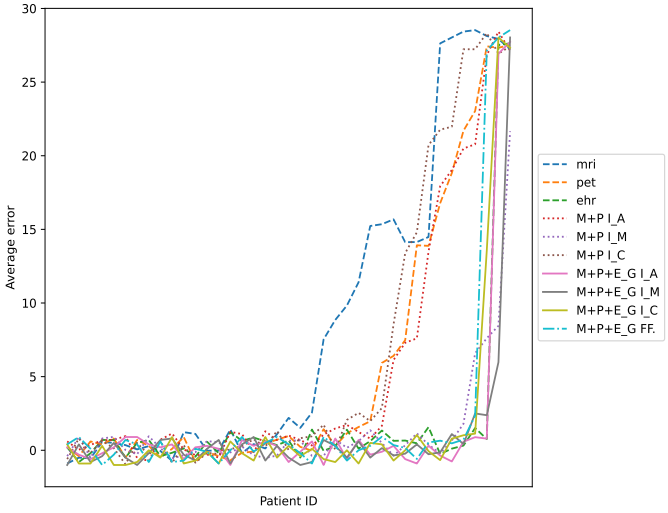
III. Results and Discussion

In this section, we compare model performance with unimodal and intermediate-level multi-modal features to demonstrate the effectiveness of our proposed method. Table III reported the evaluation metrics for test set. The combination of MRI, PET, EHRs, and genomics features (MRI+PET+EHRs+genomics) with multiplicative intermediate fusion performed the best, with an overall accuracy of 66.2% on the held-out test set. In addition, MRI+PET+EHRs+genomics with additive intermediate fusion achieved an overall accuracy of 62.7% on the test set. Lastly, MRI+PET+EHRs+genomics with concatenation intermediate fusion achieved an overall accuracy of 62.6% on the test set. As anticipated, all intermediate-level multi-modal integration methods outperformed unimodal methods.

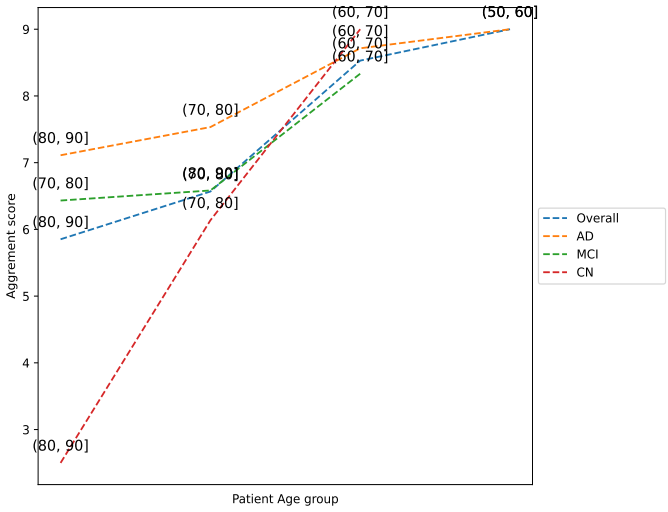
Although no singular method universally excels across all metrics, it is evident that leveraging all four modalities yields superior results compared to using only one or two modalities. Another discernible trend in our analysis is that the model performance appears to follow the sequence: Concatenation $<$ Additive $<$ Multiplicative fusion. This pattern persists in both MRI+PET and MRI+PET+EHRs+genomics scenarios.



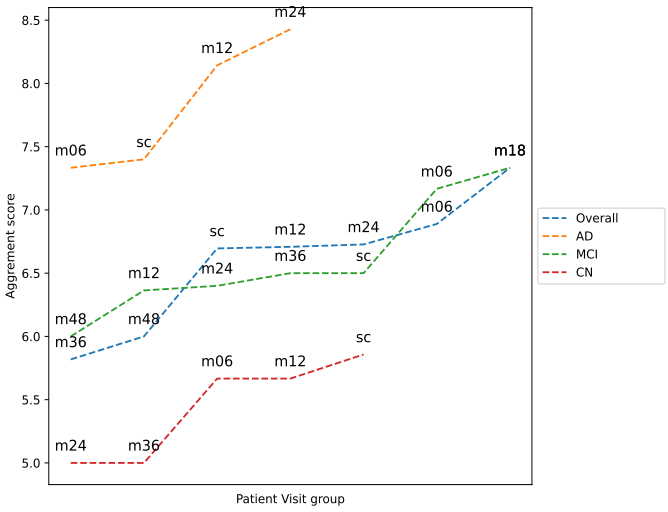
(a) Patient Age based error breakdown across models



(b) Patient group based error breakdown across models

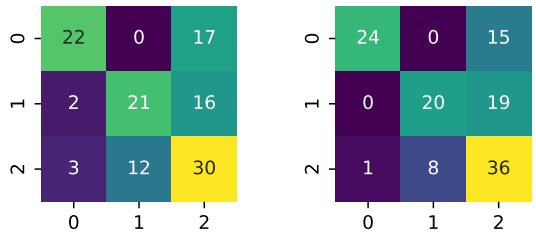


(c) Patient Age based agreement across models

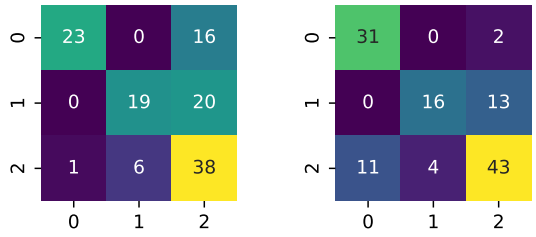


(d) Patient visit based agreement across models

Fig. 7: Error breakdown an agreement based on patient parameters. (a) Patient age-based error distribution: across all model configurations, the advanced age group of 80-90 years contributed most substantially to the error. (b) Patient group-based error distribution: The graph exhibits an exponential shape, with certain patient groups disproportionately contributing to the majority of errors. This suggests the existence of individual-specific patterns that diverge from others. (c) Patient age-based agreement: While patients within the age group of 80-90 represented the largest contributors to error, this cohort was also the subject of the greatest consensus across models (i.e., most models incorrectly agreed on predictions for this group). The 50-60 age group was the least agreed upon, indicating that performance gains were attributable to the accurate detection learned in this earlier age group. (d) Patient visit-based agreement.



(a) Additive fusion (b) Multiplicative fusion



(c) Feature concatenation (d) Feature fusion

Fig. 8: Confusion matrix for different fusion methods using MRI+PET+EHRs+genomics on the hold-out test set.

An intriguing observation is that the combination of MRI+PET performs either on par or worse than using MRI or PET in isolation. A plausible explanation for this phenomenon might be the amplified complexity of data which does not yield a commensurate increase in information. However, when incorporating MRI+PET+EHRs+genomics, the additional EHR and genomics data might supply further information that assists in interpreting the augmented complexity. As expected, the task of distinguishing between CN individuals and those with MCI, which presents the most significant challenge, registers the lowest one-vs-one accuracy across all models.

Sathvik et al [10], and Qiu et al [14] achieve an overall accuracy of 63% and 77% respectively. Junhao et al [11] achieves an AD vs CN accuracy of 86%, while Li et al [12] achieved a CN vs MCI accuracy of 74%. Our feature level fusion model achieves comparable performance with overall, AD vs CN and CN vs MCI accuracy of 73.0%, 85.1%, and 78.1% respectively. Upon examining the distribution of errors, we discovered that a majority of the errors happened in predictions of a small percentage of patients, as illustrated in Appendix Fig. 7b. In an age-wise comparison, we observed that all fusion methods exhibited superior performance in the earlier age brackets (50-70), with performance diminishing as the age group advanced (70-90). The 80-90 age group was consistently the highest contributor to errors across all methodologies (see Appendix Fig. 7a). The incorporation of age as an explicit factor during training could potentially mitigate this issue.

IV. Conclusion

In conclusion, we propose an effective end-to-end deep learning architecture designed to facilitate joint learning from MRI, PET, EHRs, and genomics data. We conduct comprehensive experiments, exploring both feature-level and intermediate-level fusion methodologies. Our results indicate that intermediate multiplicative fusion delivers superior stage prediction performance on the external validation dataset. The proposed integration framework harbors the potential to enrich our comprehension of disease diagnosis and progression by capitalizing on the complementary information derived from multimodal patient data. Moreover, the proposed multimodal deep feature integration framework may generalize to other complicated disease diagnoses and staging in real-world clinical research and practice.

References

[1] Alzheimer's Association. (2021). 2021 Alzheimer's disease facts and figures. Alzheimer's & Dementia, 17(3), 327-406.

[2] W. Wong, "Economic burden of Alzheimer disease and managed care considerations.," Am. J. Manag. Care, vol. 26, no. 8 Suppl, pp. S177–S183, Aug. 2020, doi: 10.37765/ajmc.2020.88482.

[3] M. Vaz and S. Silvestre, "Alzheimer's disease: Recent treatment strategies.," Eur. J. Pharmacol., vol. 887, p. 173554, Nov. 2020, doi: 10.1016/j.ejphar.2020.173554.

[4] M. P. Laakso et al., "MRI of the hippocampus in Alzheimer's disease: sensitivity, specificity, and analysis of the incorrectly classified subjects.," Neurobiol. Aging, vol. 19, no. 1, pp. 23–31, 1998, doi: 10.1016/s0197-4580(98)00006-2.

[5] Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu)

[6] Venugopalan, J., Tong, L., Hassanzadeh, H. R. & Wang, M. D. Multimodal deep learning models for early detection of Alzheimer's disease stage. Sci. Rep. 11, 3254 (2021).

[7] Bae, J. B. et al. Identification of Alzheimer's disease using a convolutional neural network model based on T1-weighted magnetic resonance imaging. Sci. Rep. 10, 22252 (2020).

[8] El-Sappagh, S., Alonso, J. M., Islam, S. M. R., Sultan, A. M. & Kwak, K. S. A multilayer multimodal detection and prediction model based on explainable artificial intelligence for Alzheimer's disease. Sci. Rep. 11, 2660 (2021).

[9] Lu, D. et al. Multimodal and Multiscale Deep Neural Networks for the Early Diagnosis of Alzheimer's Disease using structural MR and FDG-PET images. Sci. Rep. 8, 5697 (2018)

[10] Prabhu, Sathvik S., John A. Berkebile, Neha Rajagopalan, Renjie Yao, Wenqi Shi, Felipe Giuste, Yishan Zhong, Jimin Sun, and May D. Wang. "Multi-Modal Deep Learning Models for Alzheimer's Disease Prediction Using MRI and EHR." In 2022 IEEE 22nd International Conference on Bioinformatics and Bioengineering (BIBE), pp. 168-173. IEEE, 2022.

[11] Wen, Junhao, Jorge Samper-González, Simona Bottani, Alexandre Routier, Ninon Burgos, Thomas Jacquemont, Sabrina Fontanella et al. "Reproducible evaluation of diffusion MRI features for automatic classification of patients with Alzheimer's disease." Neuroinformatics 19 (2021): 57-78.

[12] Li, Fan, Manhua Liu, and Alzheimer's Disease Neuroimaging Initiative. "Alzheimer's disease diagnosis based on multiple cluster dense convolutional networks." Computerized Medical Imaging and Graphics 70 (2018): 101-110.

[13] Wang, Hongfei, Yanyan Shen, Shuqiang Wang, Tengfei Xiao, Liming Deng, Xiangyu Wang, and Xinyan Zhao. "Ensemble of 3D densely connected convolutional network for diagnosis of mild cognitive impairment and Alzheimer's disease." Neurocomputing 333 (2019): 145-156.

[14] Qiu, Shangran, Matthew I. Miller, Prajakta S. Joshi, Joyce C. Lee, Chonghua Xue, Yunruo Ni, Yuwei Wang et al. "Multimodal deep learning for Alzheimer's disease dementia assessment." Nature communications 13, no. 1 (2022): 3404.

[15] Sharma, A., Vans, E., Shigemizu, D., Boroevich, K. A. & Tsunoda, T. DeepInsight: A methodology to transform a non-image data to an image for convolution neural network architecture. Sci. Rep. 9, 11399 (2019).