

Deep Feature Fusion Framework for Alzheimer’s Disease Staging using Neuroimaging Modalities

Aya Gamal¹[0000–0002–8175–4554], Mustafa Elattar^{1,2}[0000–0001–7936–3522], and Sahar Selim^{1,2}[0000–0002–9886–1364]

¹ Medical Imaging and Image Processing Research Group, Center for Informatics Science, Nile University, Giza, Egypt.

² School of Information Technology and Computer Science, Nile University, Giza, Egypt.

Abstract. Alzheimer’s Disease (AD) is a significant neurodegenerative disorder. Detecting AD early is essential for effective management and improving the quality of life for both patients and their families. Recent advancements in medical imaging technology have introduced neuroimaging-based methods for early AD diagnosis. However, the challenges in early AD detection suggest that using a single modality dataset in deep learning (DL) studies, particularly neuroimaging, might not yield precise predictions of AD progression compared to integrating data from multiple imaging modalities. Utilizing information from multi-modal data fusion can enhance the detection of subtle changes and biomarkers, leading to more reliable and accurate diagnosis. In our study, we develop an automated multimodal system that integrates MRI and PET images at an intermediate fusion level, facilitating the early diagnosis of Alzheimer’s disease. This fusion approach eliminates the need for extensive preprocessing steps that are typically required in image fusion methods. Our proposed methodology outperforms previous studies in differentiating between individuals with Alzheimer’s disease and cognitively normal (CN) individuals, achieving an AUC score of 97.67% and an accuracy (ACC) of 95.24%.

Keywords: Alzheimer’s Disease · Neuroimaging Features · 3D Image Classification.

1 Introduction

Alzheimer’s Disease (AD) is a severe neurodegenerative disease. Early identification of AD is crucial for effective management and enhancement of the quality of life of both patients and their families. Unfortunately, most existing diagnostic techniques rely on subjective assessments of behavioral and cognitive symptoms, leading to potential unreliability and misdiagnosis. In recent years, advances in medical imaging technology have led to the emergence of neuroimaging-based methods for the early diagnosis of AD. However, these methods often rely on analyzing a single modality, which may fail to capture the full complexity of the

disease. Multimodal data fusion has been proposed as a promising approach to address this limitation by combining the information from different modalities.

In a clinical setting, AD is typically diagnosed by systematically examining various aspects of a patient’s multiple modalities [23]. These aspects are commonly derived from the diverse information sources of patients, including neuroimaging data, gene sequence data, profile data, and clinical mental state scale data. In contrast to the classification of AD based solely on single-modal neuroimaging, enhanced performance can be attained through the utilization of multimodal classification, involving the integration of diverse information sources. Investigating the synergies among various multimodal neuroimaging modalities significantly contributes to the identification of pathological processes in neurological disorders. This technique has been applied in image classification [25, 29] and image registration [10]. The motivation for engaging in multimodal fusion stems from two primary advantages: first, the potential for more robust predictions through the observation of the same phenomenon across multiple modalities [5]; and second, the extraction of complementary information from diverse modalities to enhance the precision of classification results [4].

The multimodal framework comprises essential components that are primarily structured at three key levels. The initial level, known as the integration level, involves defining various modalities of data intended for fusion. Thus, at this stage, a determination is made regarding what should be fused. The subsequent level is the fusion methodology, encompassing the approach employed to combine the identified data guided by the chosen fusion strategy. In the literature, fusion strategies are classified into three groups: early fusion, also known as feature-level fusion, is the process of merging multimodal data by concatenating its features into a vector, which is subsequently inputted into a machine learning model. Intermediate fusion that integrates feature representations gained from one modality at the intermediate layers of a neural network with feature representations learned from other modalities is referred to as joint fusion. Late fusion involves decision-level fusion, in which a distinct model is trained for each modality and the predictions of all models are subsequently integrated to create a final decision. The final level in the framework is the knowledge level, where the final results of the diagnosis are obtained.

Numerous studies have focused on the fusion of diverse modalities for AD diagnosis. Notably, Dwivedi et al. [9], Dong et al. [8], Xu [26], Ning [19], Hao [12], and Zhang [27] have introduced methodologies primarily focused on neuroimaging features, particularly utilizing MRI and PET modalities. Similarly, Khvostikov et al. [14], Kang [13], and Aderghal et al. [2] have directed their attention to the fusion of neuroimaging data, specifically from sMRI and DTI scans. In addition to these, Zuo et al. [30] integrated sMRI, PET, and fMRI data, while Choi and Jin [7] used fludeoxyglucose and florbetapir PET. Peng et al. [21] combined sMRI, PET, and genetic data, and Lee et al. [17] integrated cognitive performance, demographic information, CSF, and MRI imaging data.

In reviewing these studies, it is evident that the most frequently fused modalities are MRI and PET, indicating their prominent role in multimodal investi-

gations within this research domain. Various approaches to fusing MRI and PET volumes have been explored in the literature. For example, Song et al. [22] introduced a framework for AD diagnosis using a feature-fusion approach to extract semantic information from 3D MRI and PET volumes. They also proposed an image fusion method that outperformed their initial approach by reducing the number of model parameters using a single composite image, although it required multistep preprocessing. Castellano et al. [6] developed a dual branch, multimodal diagnostic model for Alzheimer’s Disease using 3D MRI and amyloid PET scans in parallel, demonstrating that these modalities provide complementary insights that enhance predictive accuracy. However, limitations include the selection of only 50 slices from the axial plane, potentially missing comprehensive spatial information, and a loss of temporal resolution in PET scans due to frame averaging.

Kong et al. [16] similarly employed an image fusion technique, while Venugopalan et al. [24] utilized 3D CNNs to extract features from MRI and PET data, demonstrating improved performance over traditional fusion methods despite being limited by dataset sizes. In contrast to CNN-based methods, transformers leverage the self-attention mechanism to capture long-range dependencies within multimodal features. Zhang et al. [28] introduced a model comprising three components: dual 3D CNN encoders for MRI and PET modalities, a Multimodal Transformer Encoder, and a classification head. They employed a feature fusion strategy that utilized a transformer-based cross-attention mechanism to fuse features more effectively. Furthermore, Miao et al. [18] utilized a transformer-based approach for multimodal multiscale fusion networks for the diagnosis of AD by fusing neuroimaging data.

In this study, we developed an automated multimodal system that integrates MRI and PET images at an intermediate fusion level, facilitating the early diagnosis of AD. This fusion method requires minimal preprocessing compared to traditional image-fusion techniques. Our approach surpasses previous studies in distinguishing between AD and CN individuals.

2 Methodology

To preserve modality-specific information for both modalities, we introduced a heuristic intermediate feature fusion framework that can capture complementary information from PET and MRI modalities independently. The components of the proposed feature fusion framework are shown in Figure 1., which illustrates the intermediate feature fusion approach that preserves modality-specific information while enabling the effective integration of MRI and PET features for classification. The first level of our framework identifies the modalities to be integrated. We then applied preprocessing steps from a streamlined pipeline to MRI and PET scans separately, preparing the data for feature extraction. In the feature extraction step, a 3D pre-trained deep learning model was used as a feature extractor for each modality. Subsequently, we employed an intermediate feature fusion approach by leveraging the feature maps extracted from the pre-

vious step and processing them for input into the classification network. Finally, a small and simple 3D CNN network was used as a classification network for the effective classification of the AD stages.

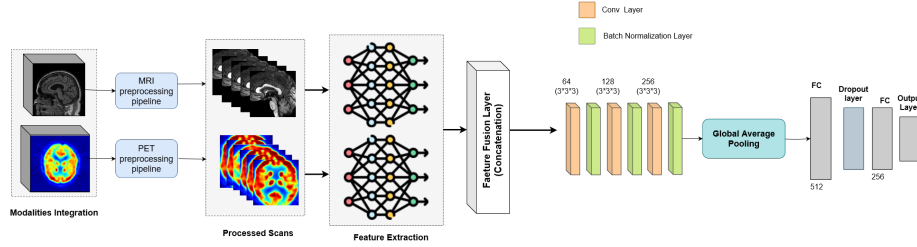


Fig. 1. The proposed intermediate feature fusion framework, highlights the stages of modality integration, feature extraction, and classification.

2.1 Dataset

Our study concentrated on the ADNI dataset (adni.loni.usc.edu), which is widely used to address this problem. We specifically implemented our experiments using structural MRI and 18-fluorodeoxyglucose (FDG)-PET modalities, which are commonly employed noninvasive methods for capturing the characteristics of brain tissue. We collected 3D data from subjects who underwent scans using both of these modalities.

First, we filtered the participants to include only those with data available on both PET and MRI during the same visit and scanning period. In total, 253 subjects participated in this experiment, contributing to a dataset of 822 scans. We aimed to mitigate the risk of data leakage by considering only the first or baseline scans for each subject. This decision ensured an equal number of scans and participants. However, to address the challenge of a small dataset size owing to the constraint of scans from the same time period, each subject could have three to four visits in different years or at least a 6-month gap within the same year. To maintain our principle of avoiding data leakage, we carefully split the data into 80%, 10%, and 10% for the training, validation, and test sets, respectively. Ensuring that a subject’s scans do not appear in different sets but all in one place.

The summary of subjects and scans in the dataset is provided in Table 1. gives an overview of the participants included in the study, highlighting the distribution across AD, MCI, and CN groups, which is crucial for understanding the dataset’s composition.

2.2 Data Preprocessing

Both MRI and FDG-PET images in ADNI underwent various processing stages. Each modality was pre-processed separately. Specifically, the MRI images underwent a series of processing steps, including skull stripping, intensity normalization, uniform resampling to achieve isotropic resolution, 3D cropping to extract

Table 1. Summary of participant statistics in the ADNI dataset (MRI and PET).

Class	Subjects	Scans
AD	43	117
MCI	111	433
CN	99	272
Total Number of scans = 822		

only the brain from the black background, resizing all scans to $128 \times 128 \times 128$, and the application of histogram equalization to enhance the contrast. The pre-processing pipeline proposed in [11] was applied here, except for the histogram equalization step, which was applied to the scans to enhance the quality and discriminatory power of the images.

Regarding PET scans preprocessing, the initial FDG-PET scans underwent the following processing steps to ensure consistency in PET data across various systems. First, we converted all the PET files into Neuroimaging Informatics Technology Initiative (NIFTI) format files, as all the processed PET image data were in the DICOM format. The dicom2nifti Python package was used to apply the conversion. Similar to MRI, PET images include extensive background regions characterized by zero pixel values beyond the brain tissue. We effectively reduced these non-essential background regions to decrease the volume of the input data via 3D cropping, as in the MRI pipeline. Furthermore, we resized the volume to $128 \times 128 \times 128$. Finally, histogram equalization was applied to the PET scans. Figure 2 outlines the PET image processing pipeline, which includes essential steps for standardizing PET images and enhancing their quality for the feature extraction process.

In this study, handling multimodal data posed a significant challenge owing to the limitations of a small sample size. To address this concern, an essential component of the proposed methodology is the augmentation step. We employed various 3D transformations on both MRI and PET data, including 3D random rotation and flipping.

2.3 Networks Architecture

The effectiveness of the 3D CNN models and transfer learning approach in diagnosing AD led us to choose them as the optimal starting point for designing our multimodal framework. The proposed multimodal model architecture is illustrated in Figure 1. The 3D DenseNet201-based transfer learning model was used as a deep feature extractor for the processed images of both modalities. The feature encoder had four dense blocks, and transition layers were employed between them. After extracting the feature maps from each modality, a concatenation layer was added to the model to fuse the intermediate features and prepare them for the single final network. The last layers in our network form a small and simple 3D CNN. The layer details of the final classification network are shown in Figure 1.

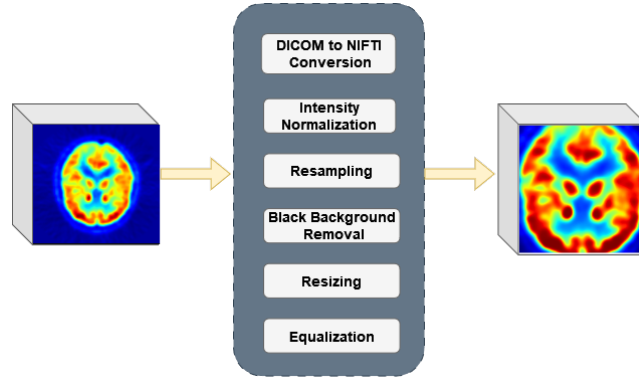


Fig. 2. PET Image Processing Method.

3 Experiments Setup and Results

In this part of our study, our experiments are organized as follows. Initially, two 3D DenseNet201 models were utilized as feature extractors for both the MRI and PET images. Subsequently, we loaded the weights of each modality independently and incorporated them into the fusion phase. Finally, a straight-forward 3D CNN network is applied to the fused features for AD diagnosis. We conducted three classification tasks: AD vs CN, AD vs MCI, and MCI vs CN.

One of the challenges highlighted in the literature is the variability in hyperparameter choices across different studies and experiments. To address this issue, we employ an open-source hyperparameter optimization framework called Optuna [3]. Optuna is compatible with any machine learning or deep learning framework, offering versatility.

We used Optuna’s automated hyperparameter optimization algorithms to efficiently explore and evaluate different configurations, facilitating the discovery of optimal model settings. Specifically, we specified the search space for the hyperparameter batch size, learning rate, and input shape by defining their types as categorical, float, and categorical with possible ranges of [5, 8, 16, 32], [0.000001, 0.0001], and [64, 96, 128], respectively. Following the optimization process, Optuna returns the best set of hyperparameters that leads to optimal performance according to the defined objective function.

The optimal configuration obtained was [batch size of 16, an input size of 128, and a learning rate of 3.4885205571560794e-05], achieved in trial 9. All experiments were performed using the TensorFlow deep learning framework [1] in Python. In the training phase of the feature extractors, we employed 200 epochs with a batch size of 16, aligned with the recommendations derived from the Optuna optimization process. Nevertheless, when training the final CNN network, we encountered hardware constraints, compelling us to decrease the batch size to five. Adam optimizer [15] is employed with a learning rate that is recommended from Optuna algorithm and a ReduceLROnPlateau strategy is

utilized here to reduce the learning rate when the validation loss has stopped improving. According to the final classification network, all the setups were the same.

In our study, we addressed the challenge of imbalanced classes by implementing oversampling and class-weighting strategies during training of our fusion model. To overcome class imbalance, we applied oversampling to the minority classes using the resample function. This step ensures that each class is adequately represented in the training dataset, thereby preventing the model from being biased towards the majority class. It randomly selects samples with replacements from the provided class indices, effectively duplicating some samples to achieve desired oversampling. This is performed until the size of the minority class matches the size of the majority class, making the class distribution more balanced in the training data. To further mitigate the impact of class imbalance, we computed class weights using the "compute class weight" function from sci-kit-learn [20]. It is used to assign different weights to different classes during model training. In the experiments, the BinaryFocalCrossentropy loss function was employed, which combines the characteristics of both binary cross-entropy (BCE) and focal loss. Binary Cross-Entropy (BCE) serves as the standard loss function for binary classification problems. On the other hand, focal loss is introduced to address class imbalance in binary classification tasks. This is achieved by modulating the cross-entropy loss and downweighting the contribution of well-classified examples where the predicted probability is high. This adjustment allowed the model to prioritize hard-to-classify examples.

We evaluated the performance of our fusion model across three binary classification tasks to recognize the three AD stages, as shown in table 2. The Table presents the performance metrics of our proposed feature fusion method, demonstrating the model's ability to differentiate between AD, MCI, and CN with high accuracy and AUC scores. Notably, the best results were achieved for the AD vs. CN task, with an AUC score of 97.67%, based on a single inference on a hold-out test set. It is important to highlight that the lower performance observed in the AD vs. MCI and MCI vs. CN tasks is expected, as the MCI stage is notoriously challenging to classify due to its overlapping features with both normal aging and early Alzheimer's, which poses a difficulty even for advanced models. Table 3 compares the performance of uni-modal approaches against our proposed feature fusion method, highlighting the substantial improvements in classification accuracy and AUC achieved through multimodal integration.

Table 2. Proposed Feature Fusion Results for 3 classification tasks.

Task	ACC	BA	AUC	F1-score
AD vs CN	95.24	95.71	97.67	93.33
MCI vs CN	80	77.81	86.08	72.86
AD vs MCI	75.0	74.23	80.54	73.4

Table 3. The uni-modal and Proposed Feature Fusion Results for AD vs CN task.

Metric	MRI	PET	Fusion method
ACC	68.75	87.5	95.24
AUC	72.5	94.29	97.67
BA	68.67	87.71	95.71
F1-score	67.15	87.67	93.33

4 Discussion and Conclusion

In this study, our goal was to utilize the power of neuroimaging multimodal data instead of unimodal data. Table 4 benchmarks the performance of our proposed method against other recent studies, showcasing the superiority of our approach in terms of accuracy across multiple classification tasks. Our proposed method outperforms other studies with superior performance for the AD vs. CN task with ACC= 95.24%.

Regarding data subjects used in this paper, instead of utilizing only the baseline scans, we obtained three to four scans for each subject in different years to overcome the small data sizes as much as possible. In addition, we took into consideration the problem of data leakage that could happen through having multiple scans for each subject, so, we split the data very carefully to ensure that the scans of each subject will not appear in different sets. By utilizing the oversampling and class weighting in our experiments, we got a superior performance of the model and we can see this effect clearly through investigating the metrics especially the f1-scores for each class in different tasks.

Integrating 3D augmentation functions significantly improved our experiments and the model’s performance. The process followed these steps: first, we applied oversampling to all training data within each classification task, balancing the minor class with the major class after splitting the data into training, validation, and test sets at the subject level. Next, preprocessing was conducted on the oversampled data. Finally, the transformation was applied exclusively to the training data, with an augmentation factor of 5.

As shown in Table 4 there are many studies, some of which follow different approaches for fusing MRI and PET volumes. Song et al. [22] introduced a framework for AD diagnosis with the feature fusion approach (intermediate fusion) to obtain semantic information from the 3D volumes of MRI and PET. In addition, they proposed another fusion method by applying an image fusion process that outperformed the first method. The image fusion approach helps reduce the number of model parameters, as a single composite image is used in the network. However, multistep pre-processing is required to achieve this fusion. Kong et al. [16] presented also an image fusion method which is considered as the early fusion approach where PET and MRI images are fused and fed into the network. In addition, Venugopalan et al. [24] suggested that the deep models for integration also showed improved performance over traditional feature-level and decision-level integrations. However, their study suffers from a limited dataset size. Zhang et al. [28] proposed an adversarial learning approach to enhance the

cross-attention mechanism for more effective feature fusion. They focused on subjects with complete T1w and FDG-PET images, utilizing feature fusion with their baseline scans. The effectiveness of their approach was then evaluated on two tasks: AD vs. CN and pMCI vs. sMCI.

Our methodology preserves modality-specific features using intermediate feature fusion, avoiding the extensive preprocessing typically required by image fusion techniques, such as volume registration and alignment. This streamlined approach led to significant improvements, particularly in the AD vs. CN task. By using a simple CNN for feature extraction and classification, our method outperformed others, highlighting the effectiveness of maintaining modality-specific information while minimizing preprocessing complexity.

The features extracted from MRI and FDG-PET scans are clinically relevant, capturing critical structural and metabolic information associated with Alzheimer’s disease. MRI highlights structural atrophy, while FDG-PET reveals metabolic changes, offering complementary insights for AD diagnosis. However, our current work lacks interpretability methods, essential for translating these features into actionable clinical insights, and faces challenges in multi-class classification due to the complexity of handling 3D data from MRI and PET scans, which requires substantial computational resources. Future work should integrate interpretability techniques to better understand the model’s decision-making process and enhance its clinical utility. Additionally, incorporating more rigorous statistical analysis, including reporting central tendencies for cross-validation runs and conducting statistical tests to confirm the significance of observed improvements, is necessary.

In conclusion, our study presented a comprehensive framework for aiding in the early diagnosis of Alzheimer’s disease through a focus on neuroimaging features. We specifically chose to focus on fusing neuroimaging features by combining 3D MRI scans with 18-FDG PET scans through the introduction of an intermediate feature fusion method. Our proposed fusion framework demonstrated superior results compared to related studies in the literature.

Table 4. Comparative performance of our classifiers and competitors.

Study	AD vs CN	MCI vs CN	AD vs MCI
	ACC (%)	ACC(%)	ACC(%)
Kong et al. (2022) [16]	93.21	86.52	85.63
song et al. (2021) [22](feature fusion)	93.22	82.37	81.00
song et al. (2021) [22](image fusion)	94.11	88.48	84.83
Venugopalan et al. (2021)[24]	86	-	-
Castellano et al. (2024)[6]	95.00	-	-
Zhang et al. (2023)[28]	92.9	-	-
Proposed feature fusion method	95.24	80	75

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Abadi, M., et al.: Tensorflow: Large-scale machine learning on heterogeneous systems (2015), <https://www.tensorflow.org/>, software available from tensorflow.org
2. Aderghal, K., Khvostikov, A., Krylov, A., Benois-Pineau, J., Karim, A., Catheline, G.: Classification of alzheimer disease on imaging modalities with deep cnns using cross-modal transfer learning. pp. 345–350 (06 2018). <https://doi.org/10.1109/CBMS.2018.00067>
3. Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M.: Optuna: A next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2019)
4. Bailey, D., Pichler, B., Gückel, B., Barthel, H., Beer, A., Bremerich, J., Czernin, J., Drzezga, A., Franzius, C., Goh, V., et al.: Combined pet/mri: multi-modality multi-parametric imaging is here: summary report of the 4th international workshop on pet/mr imaging; february 23–27, 2015, tübingen, germany. *Molecular imaging and biology* **17**, 595–608 (2015)
5. Baltrušaitis, T., Ahuja, C., Morency, L.P.: Multimodal machine learning: A survey and taxonomy. *IEEE transactions on pattern analysis and machine intelligence* **41**(2), 423–443 (2018)
6. Castellano, G., Esposito, A., Lella, E., Montanaro, G., Vessio, G.: Automated detection of alzheimer’s disease: a multi-modal approach with 3d mri and amyloid pet. *Scientific Reports* **14**(1), 5210 (2024)
7. Choi, H., Jin, K.H.: Predicting cognitive decline with deep learning of brain metabolism and amyloid imaging. *CoRR* **abs/1704.06033** (2017), <http://arxiv.org/abs/1704.06033>
8. Dong, A., Zhang, G., Liu, J., Wei, Z.: Latent feature representation learning for alzheimer’s disease classification. *Computers in Biology and Medicine* **150**, 106116 (2022). <https://doi.org/10.1016/j.compbiomed.2022.106116>
9. Dwivedi, S., Goel, T., Tanveer, M., Murugan, R., Sharma, R.: Multi-modal fusion based deep learning network for effective diagnosis of alzheimers disease. *IEEE MultiMedia* (2022)
10. Fan, J., Cao, X., Wang, Q., Yap, P.T., Shen, D.: Adversarial learning for mono-or multi-modal registration. *Medical image analysis* **58**, 101545 (2019)
11. Gamal, A., Elattar, M., Selim, S.: Automatic early diagnosis of alzheimer’s disease using 3d deep ensemble approach. *IEEE Access* **10**, 115974–115987 (2022). <https://doi.org/10.1109/access.2022.3218621>, <http://dx.doi.org/10.1109/ACCESS.2022.3218621>
12. Hao, X., Bao, Y., Guo, Y., Ming, Y., Zhang, D., Risacher, S., Saykin, A., Yao, X., Shen, L.: Multi-modal neuroimaging feature selection with consistent metric constraint for diagnosis of alzheimer’s disease. *Medical Image Analysis* **60**, 101625 (12 2019). <https://doi.org/10.1016/j.media.2019.101625>
13. Kang, L., Jiang, J., Jianjun, H., Zhang, T.: Identifying early mild cognitive impairment by multi-modality mri-based deep learning. *Frontiers in Aging Neuroscience* **12**, 206 (09 2020). <https://doi.org/10.3389/fnagi.2020.00206>
14. Khvostikov, A.V., Aderghal, K., Benois-Pineau, J., Krylov, A.S., Catheline, G.: 3d cnn-based classification using smri and md-dti images for alzheimer disease studies. *ArXiv* **abs/1801.05968** (2018), <https://api.semanticscholar.org/CorpusID:12103502>

15. Kingma, D., Ba, J.: Adam: A method for stochastic optimization. In: International Conference on Learning Representations (ICLR). San Diego, CA, USA (2015)
16. Kong, Z., Zhang, M., Zhu, W., Yi, Y., Wang, T., Zhang, B.: Multi-modal data alzheimer's disease detection based on 3d convolution. *Biomedical Signal Processing and Control* **75**, 103565 (2022)
17. Lee, G., Kang, B., Nho, K., Sohn, K.A., Kim, D.: Mildint: Deep learning-based multimodal longitudinal data integration framework. *Frontiers in Genetics* **10** (06 2019). <https://doi.org/10.3389/fgene.2019.00617>
18. Miao, S., Xu, Q., Li, W., Yang, C., Sheng, B., Liu, F., Bezabih, T.T., Yu, X.: Mmtfn: Multi-modal multi-scale transformer fusion network for alzheimer's disease diagnosis. *International Journal of Imaging Systems and Technology* **34**(1), e22970 (2024)
19. Ning, Z., Xiao, Q., Feng, Q., Chen, W., Zhang, Y.: Relation-induced multi-modal shared representation learning for alzheimer's disease diagnosis. *IEEE transactions on medical imaging* **PP** (03 2021). <https://doi.org/10.1109/TMI.2021.3063150>
20. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al.: Scikit-learn: Machine learning in python. *Journal of machine learning research* **12**(Oct), 2825–2830 (2011)
21. Peng, J., Zhu, X., An, Y., Shen, D.: Structured sparsity regularized multiple kernel learning for alzheimer's disease diagnosis. *Pattern Recognition* **88**, 370–382 (03 2019). <https://doi.org/10.1016/j.patcog.2018.11.027>
22. Song, J., Zheng, J., Li, P., Lu, X., Zhu, G., Shen, P.: An effective multimodal image fusion method using mri and pet for alzheimer's disease diagnosis. *Frontiers in Digital Health* **3**, 637386 (2021)
23. Tu, Y., Lin, S., Qiao, J., Zhuang, Y., Zhang, P.: Alzheimer's disease diagnosis via multimodal feature fusion. *Computers in Biology and Medicine* **148**, 105901 (2022)
24. Venugopalan, J., Tong, L., Hassanzadeh, H.R., Wang, M.D.: Multimodal deep learning models for early detection of alzheimer's disease stage. *Scientific reports* **11**(1), 1–13 (2021)
25. Wang, J., Wang, Q., Wang, S., Shen, D.: Sparse multi-view task-centralized learning for asd diagnosis. In: *Machine Learning in Medical Imaging: 8th International Workshop, MLMI 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, September 10, 2017, Proceedings 8*. pp. 159–167. Springer (2017)
26. Xu, H., Zhong, S., Zhang, Y.: Multi-level fusion network for mild cognitive impairment identification using multi-modal neuroimages. *Physics in medicine and biology* **68** (04 2023). <https://doi.org/10.1088/1361-6560/accac8>
27. Zhang, Y., Wang, S., Xia, K., Jiang, Y., Qian, P.: Alzheimer's disease multiclass diagnosis via multimodal neuroimaging embedding feature selection and fusion. *Information Fusion* **66**, 170–183 (02 2021). <https://doi.org/10.1016/j.inffus.2020.09.002>
28. Zhang, Y., Sun, K., Liu, Y., Shen, D.: Transformer-based multimodal fusion for early diagnosis of alzheimer's disease using structural mri and pet. In: *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*. pp. 1–5. IEEE (2023)
29. Zhou, T., Thung, K.H., Zhu, X., Shen, D.: Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis. *Human brain mapping* **40**(3), 1001–1016 (2019)
30. Zuo, Q., Lei, B., Shen, Y., Liu, Y., Feng, Z., Wang, S.Q.: Multimodal Representations Learning and Adversarial Hypergraph Fusion for Early Alzheimer's Disease Prediction, pp. 479–490 (10 2021). https://doi.org/10.1007/978-3-030-88010-1_40