Alzheimer's Disease Classification in Functional MRI With 4D Joint Temporal-Spatial Kernels in Novel 4D CNN Model [1]

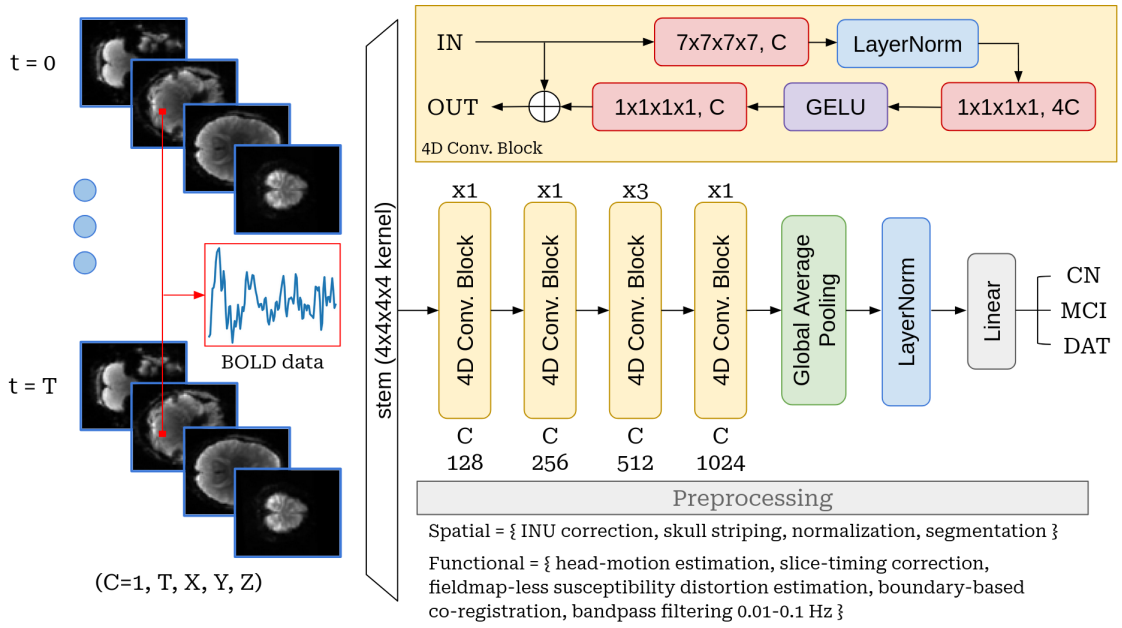

Figure 1: Proposed 4D CNN model architecture, consisting of four downsampling stages in a 1-1-3-1 configuration. The final stage outputs 1024 channels that are globally average pooled to yield 1024 features for the entire 4D scan.

Introduction

Resting-state functional MRI (rs-fMRI) is increasingly recognized as a biomarker for Alzheimer's disease (AD), with numerous studies reporting different blood-oxygen-level-dependent (BOLD) activations in specific brain regions relative to healthy subjects [2] [3].

[1] This work appeared in conference proceedings in ISMRM 2025 [1].

Feature extraction from this neuroimaging data typically involves machine learning algorithms applied to either functional connectivity matrices or subcortical surface maps [4] [5]. Contrarily, this study focuses on the 4-dimensional data for classification which is often overlooked due to bigger computational demands. Since the other approaches like timeseries data or functional connectivity matrices reduce the dimensionality of data, it is possible that some important information is lost that is beneficial for diagnosis.

We evaluate three deep learning approaches for handling 4D data. The first approach employs a 3D convolutional neural network (CNN) using the ConvNeXt [6] architecture, treating time samples as input channels. The second approach is a hybrid model combining a 3D CNN with a long short-term memory (LSTM) [7] module to separately capture spatial features and temporal dynamics. The third approach, our method, introduces a novel 4D CNN model that performs convolutions using 4D temporal-spatial kernels. While using a 4D CNN is not entirely unprecedented [8] [9], this study represents the first application of such a model in fMRI for diagnosing Alzheimer's disease.

Background

Alzheimer's Disease

Alzheimer's disease (AD) is a progressive neurodegenerative disorder that often leads to dementia, characterized by a decline in cognitive and memory functions. About 6 million Americans are currently affected by AD, and this figure is expected to rise to 12.7 million by 2050 due to the aging population [10]. Recently, three distinct stages of Alzheimer's disease have been recognized clinically: the preclinical stage, an intermediate stage known as mild cognitive impairment (MCI), and, in advanced stages, dementia of the Alzheimer's type (DAT) [11]. Patients that show no symptomps or signs of Alzheimer's disease are known to be cognitively normal (CN). MCI is diagnosed when there is objective evidence of cognitive disturbances, despite the relative preservation of daily functioning. DAT is characterized by severe cognitive and functional impairments that require clinical diagnosis. Patients with MCI, the earliest clinical phase of DAT, are at a significant risk of developing dementia, though the probability and speed of this progression vary among individuals. Additionally, preclinical AD is known to impact the brain years before any diagnosis. As a result, there is a growing need to study brain changes in the early stages to support future research on detection, prediction, and treatment strategies.

Dementia is a general term used to describe symptoms related to cognitive and memory deficits. Several diseases or disorders can lead to dementia, with Alzheimer's disease being

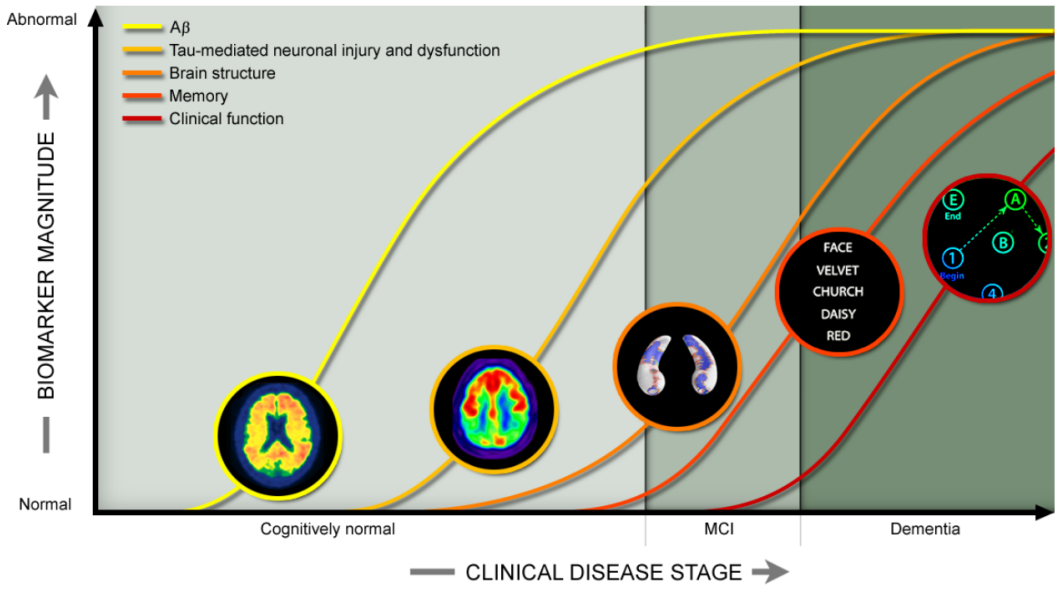

Figure 2: Schematic of Alzheimer's disease biomarkers and their progression over time. Figure from the Alzheimer's Disease Neuroimaging Initiative.

one of the most common causes. Other contributors include cerebrovascular disease, Lewy body disease, frontotemporal lobar degeneration, and Parkinson's disease, among others. However, coexisting conditions or mixed pathologies often occur between these causes, making diagnosis more complex [12]. This complexity underscores the need for adaptable biomarkers that can differentiate between or highlight overlapping pathologies when applied across different diagnoses.

Alzheimer's disease pathology is characterized by the accumulation of amyloid-β and tau proteins, which begins years or even decades before a diagnosis is made as indicated in Fig. 2. The presence of tau is associated with cognitive decline. The hippocampus, located in the temporal lobe, is one of the first brain structures affected by Alzheimer's disease. The Braak staging model describes the progression of protein accumulation, starting in the inferior temporal and medial frontal lobes and eventually affecting most areas of the brain [13]. This buildup is followed by nerve cell degeneration, or brain atrophy, which can be seen in brain imaging (such as MRI) as enlargement of the ventricles, widening of the sulci, and thinning of the gyri [14]. At this stage, cognitive and behavioral changes start to appear as the disease advances.

Cognitive changes are a natural part of normal, healthy aging, but they become more pronounced with conditions like MCI and DAT. Episodic memory, which involves the conscious recall of detailed long-term memories of unique past events, is a well-known symptom of dementia and is often associated with areas of the default-mode network [15].

However, episodic memory performance is also expected to decrease with age. On the other hand, semantic memory, which encompasses general knowledge of the world, tends to remain stable throughout life and may help differentiate between aging and pathological cognitive decline [16]. Still, it may not be consistently impaired across individuals on the Alzheimer's disease spectrum [17]. Other cognitive changes that may arise with age or dementia include declines in spatial abilities, reasoning, and processing speed. There is growing interest in studying the distinction between normal aging and pathology because Alzheimer's disease causes brain changes many years before clinical symptoms appear, and identifying early functional brain changes from normal aging can aid in early detection.

Biomarkers for Alzheimer's disease primarily target three key components: amyloid-β, tau pathology, and neuronal injury [18] [19]. To observe amyloid and tau in vivo, cerebrospinal fluid (CSF) measures and positron emission tomography (PET) are used. PET ligands, such as fluorodeoxyglucose (FDG), have become established biomarkers for Alzheimer's. FDG is used to evaluate glucose metabolism in the brain, which is often irregular in AD [20]. FDG-PET typically shows reduced metabolism in regions with brain atrophy [21] [22]. Additionally, FDG and other PET ligands are employed to study tau pathology in living patients. Neuronal injury is evaluated using structural MRI, evidence of hypometabolism in FDG-PET scans, or by measuring total tau in CSF. Structural MRI has detected AD-related brain atrophy up to ten years before the onset of symptoms and a formal diagnosis [23]. Alzheimer's disease is suspected when biomarker data and clinical cognitive evaluations point to AD pathology, which can be confirmed post-mortem through neuropathological studies using Pittsburgh Compound-B to identify amyloid-β deposits in the brain [24]. There is a clinical demand for new biomarkers to assess different aspects of Alzheimer's and to detect brain changes at earlier stages.

Increasing knowledge about the genetic aspects of Alzheimer's disease has emerged. The initial finding in this area was the link between the apolipoprotein E $\epsilon 4$ allele and an increased risk of developing Alzheimer's disease [25]. Further research has identified other genetic loci associated with the condition [26] [27], and there are notable correlations between genetic risk scores and the future risk of developing Alzheimer's, as well as the progression from MCI to DAT [28].

Increasing evidence supports the consensus that interventions should target the earliest stages of Alzheimer's disease. This underscores the need for advanced data acquisition and analysis methods to identify early brain changes, aiding in the discovery of biomarkers and the enhancement of disease detection and prediction. Research can utilize imaging techniques, cognitive assessments, and other tools to explore the structural and functional alterations associated with Alzheimer's. Functional MRI is particularly promising due to

its noninvasive nature and its ability to integrate with other modalities, such as structural MRI and Diffusion Tensor Imaging (DTI).

fMRI Biomarker in Alzheimer's Disease

Resting-state fMRI functional connectivity has been employed in numerous studies on Alzheimer's disease. The hippocampus, an area affected in the early stages of Alzheimer's, has shown disrupted resting-state functional connectivity in individuals with amnestic MCI [29]. Additionally, disrupted functional connectivity in the default mode network (DMN) is commonly observed in groups along the Alzheimer's disease spectrum. Changes in DMN connectivity have been noted in MCI and DAT patients compared to healthy controls [30] [31] [32] [33]. Impaired memory function is frequently associated with disrupted DMN functional connectivity [34] [35], and individuals with DAT exhibit reduced connectivity in the posterior DMN compared to healthy older adults [36].

Functional connectivity reveals significant differences between individuals with MCI and CN individuals, indicating that widespread degradation of brain networks can be detected in the early stages (MCI) [37]. In cases of amnestic MCI, researchers have noted initial increases in connectivity within the posterior cingulate cortex (PCC), which are followed by decreased PCC activity and heightened connectivity in the frontal network over time [38]. This pattern may indicate initial hyperactivation as a compensatory mechanism in the early stages of the disease, transitioning to hypoactivation as the pathology progresses. This progression often mirrors the pathological changes, beginning in the medial temporal lobe, spreading through regions of the DMN, including the PCC, and eventually reaching the frontal areas of the brain in later stages [39].

Understanding the neural mechanisms involved in early brain changes associated with Alzheimer's disease is vital for predicting the progression to advanced stages and researching therapeutic interventions. The fact that functional connectivity is a sensitive indicator of memory and other Alzheimer related changes in the brain, coupled with evidence of increased functional connectivity in the initial stages, suggests that it could be useful for early detection and may offer additional benefits for prediction studies.

Neuroimaging has become a crucial tool in the clinical assessment of individuals suspected of having neurodegenerative diseases like Alzheimer's. Structural MRI brain scans can reveal the presence and progression of neurodegeneration. Individuals with MCI and DAT often show the characteristic progressive atrophy associated with AD, particularly in regions like the medial temporal lobes [40]. However, by the time significant brain atrophy is detected, the disease may have already been affecting the brain for years or even

decades, making early detection essential. Current research focuses on better characterizing the early functional changes in the brain linked to AD, with functional MRI being one approach, as it serves as a proxy for neural activity.

Functional MRI measures are increasingly being studied as potential biomarkers for Alzheimer's disease, with a focus on connectivity and network analysis. As noted, compensatory increases in fMRI activations during the early stages of Alzheimer's pathology are widely recognized, while general reductions in brain activity are often observed during advanced stages. Differences in fMRI activations might be associated with altered neural activity leading to impairments, such as memory deficits, or with neurovascular dysfunction affecting neurovascular coupling, among other possible reasons. The key takeaway is that fMRI can provide an indirect assessment of neuronal functioning and may help identify patients at risk of developing AD before significant atrophy occurs. Thus, detection and interventions are critical at the earliest stages of the disease, and fMRI is a promising tool due to its noninvasive nature and ability to integrate with other modalities, potentially illuminating neural mechanisms involved in early changes and leading to improved early detection.

Long Short-Term Memory (LSTM)

A Long Short-Term Memory (LSTM) [7] is a type of recurrent neural network (RNN) architecture that is particularly well-suited for modeling sequential data and capturing long-range dependencies. The LSTM addresses the issue of vanishing gradients, which can arise during the training of traditional RNNs. It does this by introducing a more complex, gated architecture. LSTMs are powerful due to their ability to maintain information over prolonged sequences using the cell state and multiple gates to control information flow. As a consequence, they are commonly employed in areas such as language modeling, time-series prediction, and any other application where capturing temporal dependencies is critical. The architecture for an LSTM layer is described below.

An LSTM cell comprises several components designed to regulate the flow of information, namely: cell state C_t, hidden state h_t, and gates f_t, i_t, o_t. The cell state C_t is the internal memory of the LSTM cell that carries information across different time steps. It is modulated by various gates to retain essential information over long periods. The hidden state h_t is the output of the LSTM cell at time step t, which is used both as output and as input to other model components at the next time step. The LSTM cell contains three key gates. The first one, forget gate f_t, determines which information from the cell state should be discarded. It uses a sigmoid activation function to produce values between 0 and

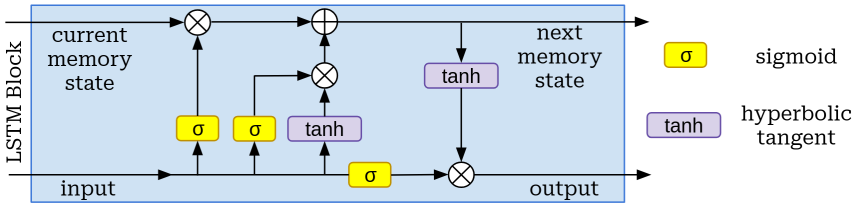

Figure 3: Illustration of LSTM layer used in deep learning models.

1, which are then applied element-wise to the cell state, written as

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f). \tag{1}$$

The input gate i_t, decides which new information should be added to the cell state. It also uses a sigmoid function to serve as a filter for input modulation, as expressed below

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i). \tag{2}$$

The output gate o_t, determines what part of the cell state should be output as the hidden state of the current time step. This gate uses sigmoid functions as a gate mechanism before computing the new hidden state as indicated

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o). \tag{3}$$

The cell state is updated using the forget gate and the candidate updates. This layer generates potential values that could be added to the cell state, typically using a hyperbolic tangent activation function. The new cell state is computed along with the hidden state using the output gate and cell state. This is mathematically written below and illustrated in Fig. 3.

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{4}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \tag{5}$$

$$h_t = o_t \odot \tanh(C_t). \tag{6}$$

Global Average Pooling (GAP)

Global Average Pooling (GAP) [41] is a data processing technique used to distill a high-dimensional data representation into a simplified form by computing the average value

across particular dimensions of a tensor, in this context used to both temporally and spatially compress the tensor. GAP operates by aggregating information over the entirety of specified dimensions, thereby condensing a feature set into a singular representation for each channel or feature dimension. This operation effectively transforms a multi-dimensional data array into a lower-dimensional tensor, often reducing dimensions associated with spatial characteristics to a singular mean value per feature channel. Since our data is size $(B, C, S_X, S_Y, S_Z, S_T)$ for B batch size, C channels after many convolutional layers with reduced spatial and time dimensions (S_X, S_Y, S_Z, S_T), GAP will reduce the input to size (B, C) by averaging over (S_X, S_Y, S_Z, S_T) in this context. Note that (S_X, S_Y, S_Z, S_T) is very small in general at the end of any CNN, in our case $3 \times 3 \times 3 \times 4$. This process yields a compact representation, facilitating the use of a linear layer for classification. Unlike many other neural network operations that require parameter learning, GAP functions independently of learned parameters. GAP can handle input data of variable sizes because the averaging process is independent of the absolute dimensions, focusing instead on the mean value across dimensions. This property is particularly advantageous in this situation since there may be some variation in original input size depending on spatial voxel size and repetition time (sampling rate).

Layer Normalization

Layer Normalization [42] is a technique used in neural networks to improve training stability and convergence by normalizing the activations of intermediate layers. It addresses the problem of internal covariate shift, where the distribution of inputs to a given layer changes during training, potentially slowing down the learning process. Consider a layer with activations represented by a vector $\mathbf{x} = [x_1, x_2, \ldots, x_H]$, where H denotes the number of hidden units in the layer. For layer normalization, the mean and variance are computed across the layer's units

$$\mu = \frac{1}{H} \sum_{i=1}^{H} x_i \tag{7}$$

$$\sigma^2 = \frac{1}{H} \sum_{i=1}^{H} (x_i - \mu)^2. \tag{8}$$

Then, each element x_i is then normalized using the computed mean and variance:

$$\hat{x}_i = \frac{x_i - \mu}{\sqrt{\sigma^2 + \epsilon}} \tag{9}$$

where ϵ is a small constant added for numerical stability to prevent division by zero. Finally, just like batch normalization, the normalized activations are scaled and shifted using learnable parameters γ (scale) and β (shift) by

$$y_i = \gamma \hat{x}_i + \beta. \tag{10}$$

Here, γ and β are parameters that are learned during the training process, allowing the network to adapt the normalized output to the desired range of activations.

Layer normalization normalizes each data point independently across its features, unlike batch normalization, which normalizes across the batch dimension. This characteristic makes layer normalization particularly useful for recurrent neural networks or scenarios with small batch sizes. By stabilizing the distribution of inputs to each layer throughout the training process, layer normalization can lead to faster convergence and potentially improved performance. The additional computational complexity of layer normalization is relatively low, as it involves only the computation of per-layer statistics and the learning of a linear transformation.

Gaussian Error Linear Unit (GELU)

The Gaussian Error Linear Unit (GELU) [43] is a novel activation function used in neural networks, particularly in the context of deep learning. It differs from traditional activation functions like ReLU (Rectified Linear Unit) by incorporating stochastic regularization, which has been shown to improve performance in several tasks.

The GELU activation function is defined as

$$\text{GELU}(x) = x \cdot \Phi(x) \tag{11}$$

where $\Phi(x)$ is the cumulative distribution function (CDF) of the standard normal distribution. This can be approximated using either $\text{erf}(\cdot)$, the Gaussian error function, or using hyperbolic tangent as follows

$$\Phi(x) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{x}{\sqrt{2}}\right)\right] \tag{12}$$

$$\Phi(x) \approx \frac{1}{2} \cdot \left(1 + \tanh\left(\sqrt{\frac{2}{\pi}}\left(x + 0.044715 \cdot x^3\right)\right)\right) \tag{13}$$

Unlike deterministic activations such as ReLU, GELU introduces an element of stochasticity by weighing inputs based on how they compare to their own normal distribution. This

allows GELU to decide which neurons to activate in a softer manner than the hard thresholding approach of ReLU. Because of this, GELU behaves smoothly over the range of inputs by preventing issues with zero gradients (a problem encountered in ReLU with negative inputs). By employing a Gaussian-based non-linearity, GELU enables networks to process input data more fluidly across its activation spectrum, offering a compelling alternative to traditional activation functions like ReLU. This function has shown to contribute to the capacity of models for learning complex patterns in data efficiently.

4D Kernels

4D spatial-temporal convolutions are aimed at processing data that possesses both spatial and temporal dimensions. Such operations are pivotal when the data to be analyzed evolves over time, necessitating a model capable of capturing dependencies across both spatial and temporal contexts. Incorporating 4D convolutions extends traditional convolution operations to handle datasets that are represented in a $C \times X \times Y \times Z \times T$ format. Here, C denotes the number of input channels, while X, Y, and Z correspond to spatial dimensions, and T signifies the temporal dimension.

In essence, a 4D convolution layer applies a series of learnable filters (or kernels) across the input data, capturing local patterns not only spatially but also across temporal sections. This operation simultaneously processes information in these dimensions, leading to a comprehensive understanding of how spatial features change, remain consistent, or evolve through time or sequences. For instance, in video analysis, understanding the correlation and progression of scenes is essential. Additionally, in medical imaging, 4D convolutions might be employed to analyze dynamic sequences of volumetric data, such as in functional MRI scans, where spatial patterns within bodily structures need to be interpreted over time to gain further insights. Furthermore, such convolutions often lead to reductions in computational complexity when compared with separate spatial and temporal processing steps, allowing more integrative and resource-efficient learning. Thus, a 4D CNN represents a robust tool for modeling and understanding datasets with inherent multi-dimensional relationships, significantly enhancing the capacity of neural networks to handle rich and complex data forms.

Methods

We used the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset [44], selecting 3T rs-fMRI scans (MPRAGE/SPGR pulse sequence data) with spatial resolution 3.3mm

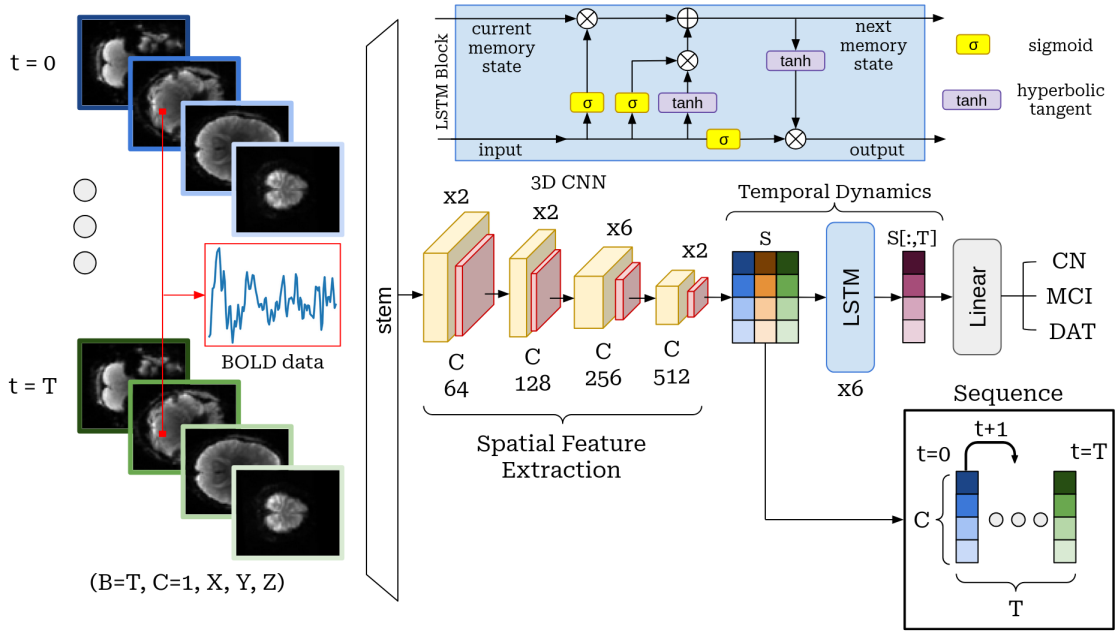

Figure 4: Architecture of the hybrid 3D CNN + LSTM model. Each 3D time sample is processed individually by the CNN, and the resulting features are aggregated into matrix S. The LSTM module captures temporal dynamics between time samples for classification purposes.

Method	Accuracy	Sensitivity	Specificity
2 class (CN/DAT)			
3D CNN	0.68	0.54	0.84
3D CNN + LSTM	0.72	0.58	0.88
4D CNN	0.77	0.62	0.92
2 class (CN/MCI)			
3D CNN	0.58	0.66	0.50
3D CNN + LSTM	0.61	0.40	0.82
4D CNN	0.63	0.64	0.62
3 class (CN/MCI/DAT)			
3D CNN	0.48	0.48	0.82
3D CNN + LSTM	0.50	0.50	0.78
4D CNN	0.53	0.53	0.83

Figure 5: Comparative results for the three approaches to handling the time dimension in raw 4D fMRI data. Accuracy, sensitivity, and specificity are reported for various class settings (binary and multi-class classification) using the ADNI test dataset.

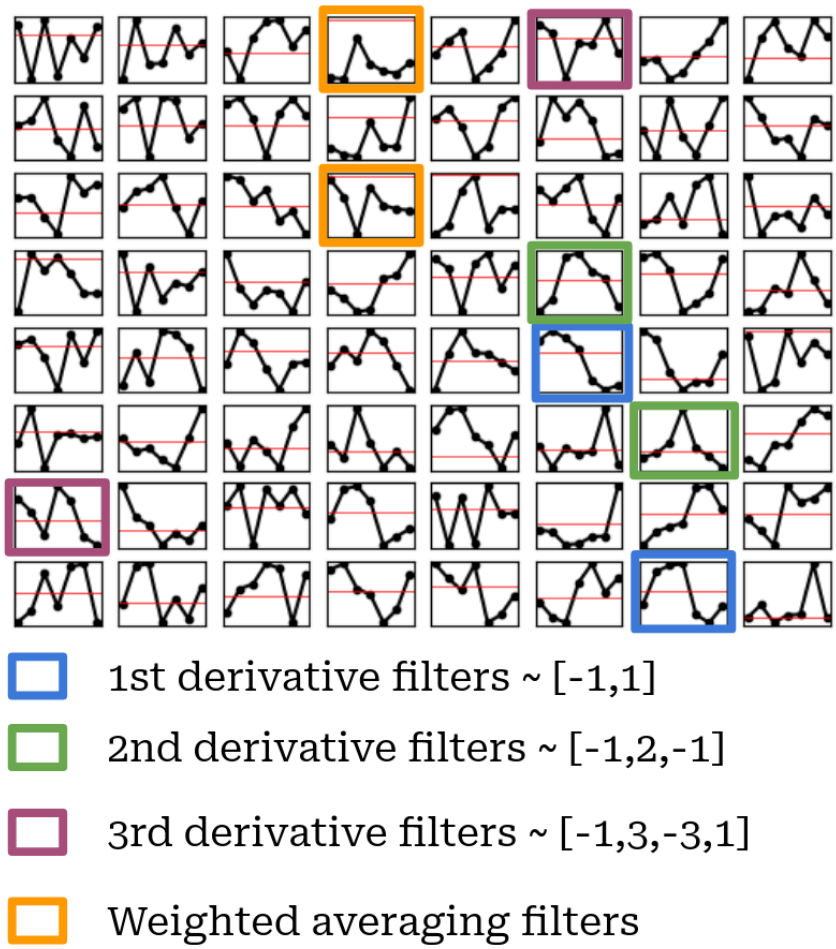

Figure 6: Temporal kernels from random spatial kernel locations for first layer channels (C=128). Only a subset of the total channels are shown for illustration simplicity. Moreover, only a few examples per filter are shown. The proposed model in the first layer extracts low-level features by using derivative and weighted average filters among other kernels less interpretable.

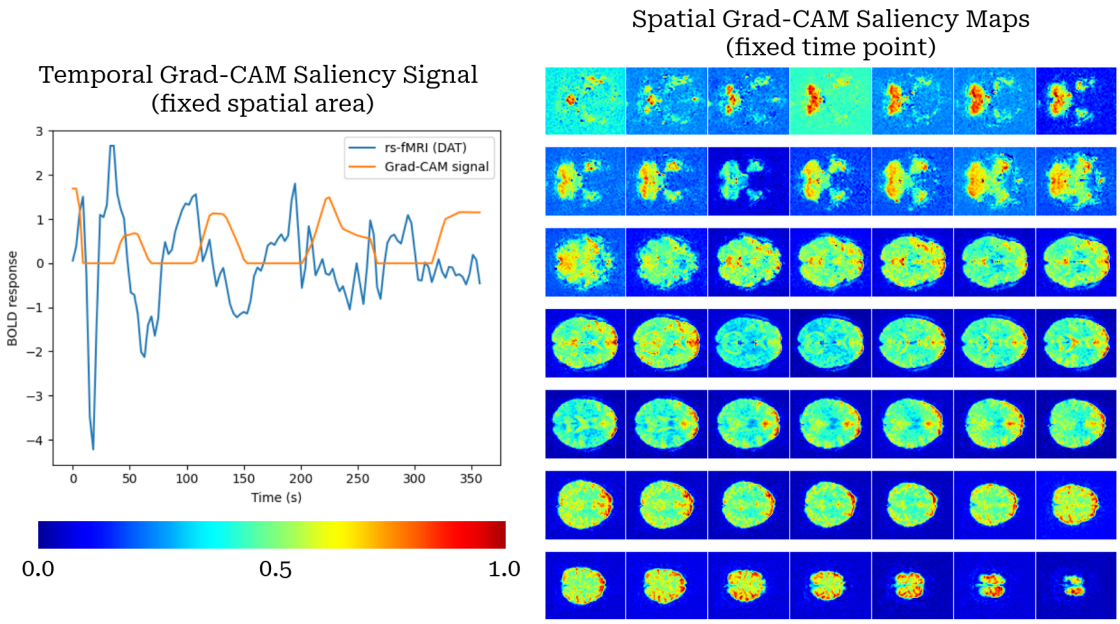

Figure 7: Model interpretability figure using the Grad-CAM++ method. Left image consists of the BOLD response at the hippocampus for a DAT diagnosed subject and corresponding Grad-CAM saliency signal over time. Right image consists of spatial Grad-CAM maps for a fixed time sample, illustrating key regions used for classification.

and a repetition time of 3000ms, comprising 140 temporal samples of size 65×77×65. Additionally, structural MRI scans were used to assist in the normalization process. Recognizing the usually limited size of medical datasets, we augmented the dataset by considering each session as an independent "pseudo-subject" to mitigate class imbalance and increase dataset size. To prevent cross-contamination between train and test sets, scans from the same individual were assigned exclusively to one set, resulting in class distributions CN (602/50), MCI (210/50), and DAT (147/50) for train/test samples. The test set was balanced to ensure accuracy is a meaningful metric and the validation set consisted of a subset of the train set through k-fold cross-validation. Data preprocessing involved several steps: converting raw DICOM files to the BIDS format [45] and processing with fmriprep [46]. For structural scans, this included N4 bias field correction, skull stripping, and spatial normalization to the MNI152 linear space from TemplateFlow [47]. For functional scans, this entailed slice-timing correction, head-motion estimation, and fieldmap-less susceptibility distortion correction. Further preprocessing involved bandpass filtering between 0.01-0.1 Hz using scikit-learn [48], discarding the first 20 temporal samples, and applying Z-score normalization on each voxel's time series data. All models were implemented using PyTorch, with data importation facilitated by the NiBabel [49] package. We addressed class imbalance using a weighted cross-entropy loss function with inverse frequency weights

w=[959/602,959/210,959/147] and used the Adam optimizer [50] with weight decay and a cosine decay learning rate scheduler. Training was conducted on a system equipped with a 32-core CPU, 187GB RAM, and 1 NVIDIA 4090 24GB-VRAM GPU. For the 4D CNN model, custom "Conv4D" layers were developed and integrated into our 4D convolutional blocks as illustrated in Fig. 1. For the 3D CNN + LSTM model, spatial features for each time sample were separately extracted and globally averaged. Then, all of the collected time samples were provided to the LSTM module to be used for classification as shown in Fig. 4. For the 3D CNN model, all time samples were treated as separate channel inputs.

Results

The 4D CNN model better predicted patient diagnosis compared to other models as indicated in Fig. 5. For model interpretability, two analyses were conducted. First, the 4D kernels in the first layer were plotted against time to visualize features learned by the model (Fig. 6). From this, it appears that some lower-level features include derivative and averaging information. Secondly, the Grad-CAM++ method [51] was employed to identify important regions used for diagnosis. Fig. 7 presents the BOLD response in the hippocampus with corresponding Grad-CAM signal, and to the right, spatial saliency maps at a fixed time point, highlighting significant spatial features such as cerebellum, prefrontal cortex, and hippocampus.

Discussion/Conclusion

This study demonstrates that diagnosis can be predicted using joint temporal-spatial kernels, illustrating the efficacy of a 4D CNN. The model outperformed other methods that rely on more conventional modeling assumptions such as separate spatial and temporal learning modules. Moreover, saliency maps indicate relevant brain regions used for diagnosis.

Future research directions can focus on extensions to task-based fMRI data, where "stressing" various networks beyond the default mode network could provide more insight into cognitive performance and Alzheimer's diagnosis. Additionally, the model could be restructured to work on regression-based tasks, such as score prediction, rather than classification.

BIBLIOGRAPHY

[1] J. Salazar Cavazos and S. Peltier, "Alzheimer's disease classification in functional mri with 4d joint temporal-spatial kernels in novel 4d cnn model," in *International Society for Magnetic Resonance in Medicine*, p. 6929, ISMRM, 2025.

[2] P. Vemuri, D. T. Jones, and C. R. Jack, "Resting state functional mri in alzheimer's disease," *Alzheimer's research & therapy*, vol. 4, pp. 1–9, 2012.

[3] R. Sperling, "The potential of functional mri as a biomarker in early alzheimer's disease," *Neurobiology of aging*, vol. 32, pp. S37–S43, 2011.

[4] P. Vemuri, D. T. Jones, and C. R. Jack, "Resting state functional mri in alzheimer's disease," *Alzheimer's research & therapy*, vol. 4, pp. 1–9, 2012.

[5] M. Karker, *Predictive Analysis and Deep Learning of Functional MRI in Alzheimer's Disease*. PhD thesis, University of Michigan, 2022.

[6] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11976–11986, 2022.

[7] Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: Lstm cells and network architectures," *Neural computation*, vol. 31, no. 7, pp. 1235–1270, 2019.

[8] A. Myronenko, D. Yang, V. Buch, D. Xu, A. Ihsani, S. Doyle, M. Michalski, N. Tenenholtz, and H. Roth, "4d cnn for semantic segmentation of cardiac volumetric sequences," in *Statistical Atlases and Computational Models of the Heart. Multi-Sequence CMR Segmentation, CRT-EPiggy and LV Full Quantification Challenges: 10th International Workshop, STACOM 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019, Revised Selected Papers 10*, pp. 72–80, Springer, 2020.

[9] M. Bengs, N. Gessert, and A. Schlaefer, "4d spatio-temporal deep learning with 4d fmri data for autism spectrum disorder classification," *arXiv preprint arXiv:2004.10165*, 2020.

[10] K. B. Rajan, J. Weuve, L. L. Barnes, E. A. McAninch, R. S. Wilson, and D. A. Evans, "Population estimate of people with clinical alzheimer's disease and mild cognitive impairment in the united states (2020–2060)," *Alzheimer's & dementia*, vol. 17, no. 12, pp. 1966–1975, 2021.

[11] C. R. Jack Jr, D. A. Bennett, K. Blennow, M. C. Carrillo, B. Dunn, S. B. Haeberlein, D. M. Holtzman, W. Jagust, F. Jessen, J. Karlawish, *et al.*, "Nia-aa research framework: toward a biological definition of alzheimer's disease," *Alzheimer's & dementia*, vol. 14, no. 4, pp. 535–562, 2018.

[12] G. D. Rabinovici, M. C. Carrillo, M. Forman, S. DeSanti, D. S. Miller, N. Kozauer, R. C. Petersen, C. Randolph, D. S. Knopman, E. E. Smith, *et al.*, "Multiple comorbid neuropathologies in the setting of alzheimer's disease neuropathology and implications for drug development," *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, vol. 3, no. 1, pp. 83–91, 2017.

[13] T. Ohm, H. Müller, H. Braak, and J. Bohl, "Close-meshed prevalence rates of different stages as a tool to uncover the rate of alzheimer's disease-related neurofibrillary changes," *Neuroscience*, vol. 64, no. 1, pp. 209–217, 1995.

[14] P. Vemuri and C. R. Jack, "Role of structural mri in alzheimer's disease," *Alzheimer's research & therapy*, vol. 2, pp. 1–10, 2010.

[15] M. D. Greicius, G. Srivastava, A. L. Reiss, and V. Menon, "Default-mode network activity distinguishes alzheimer's disease from healthy aging: evidence from functional mri," *Proceedings of the National Academy of Sciences*, vol. 101, no. 13, pp. 4637–4642, 2004.

[16] R. D. Nebes, "Semantic memory in alzheimer's disease.," *Psychological bulletin*, vol. 106, no. 3, p. 377, 1989.

[17] J. R. Hodges and K. Patterson, "Is semantic memory consistently impaired early in the course of alzheimer's disease? neuroanatomical and diagnostic implications," *Neuropsychologia*, vol. 33, no. 4, pp. 441–459, 1995.

[18] C. R. Jack Jr, D. A. Bennett, K. Blennow, M. C. Carrillo, H. H. Feldman, G. B. Frisoni, H. Hampel, W. J. Jagust, K. A. Johnson, D. S. Knopman, *et al.*, "A/t/n: An unbiased descriptive classification scheme for alzheimer disease biomarkers," *Neurology*, vol. 87, no. 5, pp. 539–547, 2016.

[19] C. R. Jack Jr, D. S. Knopman, S. D. Weigand, H. J. Wiste, P. Vemuri, V. Lowe, K. Kantarci, J. L. Gunter, M. L. Senjem, R. J. Ivnik, *et al.*, "An operational approach to national institute on aging–alzheimer's association criteria for preclinical alzheimer disease," *Annals of neurology*, vol. 71, no. 6, pp. 765–775, 2012.

[20] L. Mosconi, R. Mistur, R. Switalski, W. H. Tsui, L. Glodzik, Y. Li, E. Pirraglia, S. De Santi, B. Reisberg, T. Wisniewski, *et al.*, "Fdg-pet changes in brain glucose metabolism from normal cognition to pathologically verified alzheimer's disease," *European journal of nuclear medicine and molecular imaging*, vol. 36, pp. 811–822, 2009.

[21] M. De Leon, A. Convit, O. Wolf, C. Tarshish, S. DeSanti, H. Rusinek, W. Tsui, E. Kandil, A. Scherer, A. Roche, *et al.*, "Prediction of cognitive decline in normal

elderly subjects with 2-[18f] fluoro-2-deoxy-d-glucose/positron-emission tomography (fdg/pet)," *Proceedings of the National Academy of Sciences*, vol. 98, no. 19, pp. 10966–10971, 2001.

[22] L. Mosconi, A. Pupi, and M. J. De Leon, "Brain glucose hypometabolism and oxidative stress in preclinical alzheimer's disease," *Annals of the New York Academy of Sciences*, vol. 1147, no. 1, pp. 180–195, 2008.

[23] M. Tondelli, G. K. Wilcock, P. Nichelli, C. A. De Jager, M. Jenkinson, and G. Zamboni, "Structural mri changes detectable up to ten years before clinical alzheimer's disease," *Neurobiology of aging*, vol. 33, no. 4, pp. 825–e25, 2012.

[24] M. D. Ikonomovic, W. E. Klunk, E. E. Abrahamson, C. A. Mathis, J. C. Price, N. D. Tsopelas, B. J. Lopresti, S. Ziolko, W. Bi, W. R. Paljug, *et al.*, "Post-mortem correlates of in vivo pib-pet amyloid imaging in a typical case of alzheimer's disease," *Brain*, vol. 131, no. 6, pp. 1630–1645, 2008.

[25] J. Poirier, P. Bertrand, S. Kogan, S. Gauthier, J. Davignon, and D. Bouthillier, "Apolipoprotein e polymorphism and alzheimer's disease," *The Lancet*, vol. 342, no. 8873, pp. 697–699, 1993.

[26] J.-C. Lambert, S. Heath, G. Even, D. Campion, K. Sleegers, M. Hiltunen, O. Combarros, D. Zelenika, M. J. Bullido, B. Tavernier, *et al.*, "Genome-wide association study identifies variants at clu and cr1 associated with alzheimer's disease," *Nature genetics*, vol. 41, no. 10, pp. 1094–1099, 2009.

[27] S. Seshadri, A. L. Fitzpatrick, M. A. Ikram, A. L. DeStefano, V. Gudnason, M. Boada, J. C. Bis, A. V. Smith, M. M. Carrasquillo, J. C. Lambert, *et al.*, "Genome-wide analysis of genetic loci associated with alzheimer disease," *Jama*, vol. 303, no. 18, pp. 1832–1840, 2010.

[28] C. Bellenguez, F. Küccükali, I. E. Jansen, L. Kleineidam, S. Moreno-Grau, N. Amin, A. C. Naj, R. Campos-Martin, B. Grenier-Boley, V. Andrade, *et al.*, "New insights into the genetic etiology of alzheimer's disease and related dementias," *Nature genetics*, vol. 54, no. 4, pp. 412–436, 2022.

[29] F. Bai, C. Xie, D. R. Watson, Y. Shi, Y. Yuan, Y. Wang, C. Yue, Y. Teng, D. Wu, and Z. Zhang, "Aberrant hippocampal subregion networks associated with the classifications of amci subjects: a longitudinal resting-state study," *PloS one*, vol. 6, no. 12, p. e29288, 2011.

[30] F. Agosta, M. Pievani, C. Geroldi, M. Copetti, G. B. Frisoni, and M. Filippi, "Resting state fmri in alzheimer's disease: beyond the default mode network," *Neurobiology of aging*, vol. 33, no. 8, pp. 1564–1578, 2012.

[31] M. A. Binnewijzend, M. M. Schoonheim, E. Sanz-Arigita, A. M. Wink, W. M. van der Flier, N. Tolboom, S. M. Adriaanse, J. S. Damoiseaux, P. Scheltens, B. N. van Berckel, *et al.*, "Resting-state fmri changes in alzheimer's disease and mild cognitive impairment," *Neurobiology of aging*, vol. 33, no. 9, pp. 2018–2028, 2012.

[32] S. A. Rombouts, F. Barkhof, R. Goekoop, C. J. Stam, and P. Scheltens, "Altered resting state networks in mild cognitive impairment and mild alzheimer's disease: an fmri study," *Human brain mapping*, vol. 26, no. 4, pp. 231–239, 2005.

[33] K. Wang, M. Liang, L. Wang, L. Tian, X. Zhang, K. Li, and T. Jiang, "Altered functional connectivity in early alzheimer's disease: A resting-state fmri study," *Human brain mapping*, vol. 28, no. 10, pp. 967–978, 2007.

[34] S. M. Daselaar, S. E. Prince, and R. Cabeza, "When less means more: deactivations during encoding that predict subsequent memory," *Neuroimage*, vol. 23, no. 3, pp. 921–927, 2004.

[35] S. L. Miller, K. Celone, K. DePeau, E. Diamond, B. C. Dickerson, D. Rentz, M. Pihlajamäki, and R. A. Sperling, "Age-related memory impairment associated with loss of parietal deactivation but preserved hippocampal activation," *Proceedings of the National Academy of Sciences*, vol. 105, no. 6, pp. 2181–2186, 2008.

[36] M. D. Greicius, G. Srivastava, A. L. Reiss, and V. Menon, "Default-mode network activity distinguishes alzheimer's disease from healthy aging: evidence from functional mri," *Proceedings of the National Academy of Sciences*, vol. 101, no. 13, pp. 4637–4642, 2004.

[37] S.-Y. Lin, C.-P. Lin, T.-J. Hsieh, C.-F. Lin, S.-H. Chen, Y.-P. Chao, Y.-S. Chen, C.-C. Hsu, and L.-W. Kuo, "Multiparametric graph theoretical analysis reveals altered structural and functional network topology in alzheimer's disease," *NeuroImage: Clinical*, vol. 22, p. 101680, 2019.

[38] F. Bai, D. R. Watson, Y. Shi, Y. Wang, C. Yue, YuhuanTeng, D. Wu, Y. Yuan, and Z. Zhang, "Specifically progressive deficits of brain functional marker in amnestic type mild cognitive impairment," *PloS one*, vol. 6, no. 9, p. e24271, 2011.

[39] H. Braak, I. Alafuzoff, T. Arzberger, H. Kretzschmar, and K. Del Tredici, "Staging of alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry," *Acta neuropathologica*, vol. 112, no. 4, pp. 389–404, 2006.

[40] P. Vemuri and C. R. Jack, "Role of structural mri in alzheimer's disease," *Alzheimer's research & therapy*, vol. 2, pp. 1–10, 2010.

[41] Y. Dogan, "A new global pooling method for deep neural networks: Global average of top-k max-pooling," *Traitement du signal*, vol. 40, no. 2, pp. 577–587, 2023.

[42] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," *arXiv preprint arXiv:1607.06450*, 2016.

[43] D. Hendrycks and K. Gimpel, "Gaussian error linear units (gelus)," *arXiv preprint arXiv:1606.08415*, 2016.

[44] S. G. Mueller, M. W. Weiner, L. J. Thal, R. C. Petersen, C. Jack, W. Jagust, J. Q. Trojanowski, A. W. Toga, and L. Beckett, "The alzheimer's disease neuroimaging initiative," *Neuroimaging Clinics of North America*, vol. 15, no. 4, p. 869, 2005.

[45] K. J. Gorgolewski, T. Auer, V. D. Calhoun, R. C. Craddock, S. Das, E. P. Duff, G. Flandin, S. S. Ghosh, T. Glatard, Y. O. Halchenko, *et al.*, "The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments," *Scientific data*, vol. 3, no. 1, pp. 1–9, 2016.

[46] O. Esteban, C. J. Markiewicz, R. W. Blair, C. A. Moodie, A. I. Isik, A. Erramuzpe, J. D. Kent, M. Goncalves, E. DuPre, M. Snyder, *et al.*, "fmriprep: a robust preprocessing pipeline for functional mri," *Nature methods*, vol. 16, no. 1, pp. 111–116, 2019.

[47] R. Ciric, W. H. Thompson, R. Lorenz, M. Goncalves, E. E. MacNicol, C. J. Markiewicz, Y. O. Halchenko, S. S. Ghosh, K. J. Gorgolewski, R. A. Poldrack, *et al.*, "Templateflow: Fair-sharing of multi-scale, multi-species brain models," *Nature Methods*, vol. 19, no. 12, pp. 1568–1571, 2022.

[48] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, *et al.*, "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

[49] M. Brett, C. J. Markiewicz, M. Hanke, M.-A. Côté, B. Cipollini, P. McCarthy, D. Jarecka, C. P. Cheng, Y. O. Halchenko, M. Cottaar, *et al.*, "nipy/nibabel: 3.2. 1," *Zenodo*, 2020.

[50] D. P. Kingma, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[51] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, pp. 618–626, 2017.