

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

Javier Salazar Cavazos<sup>1</sup> and Scott Peltier<sup>2</sup>

<sup>1</sup>*Electrical & Computer Engineering, University of Michigan, Ann Arbor, MI, United States*

<sup>2</sup>*Biomedical Engineering, University of Michigan, Ann Arbor, MI, United States*

## Synopsis

**Keywords:** Diagnosis/Prediction, Analysis/Processing

**Motivation:** Previous works in the literature apply 3D spatial-only models on 4D functional MRI data leading to possible sub-par feature extraction to be used for downstream tasks like classification.

**Goal(s):** In this work, we aim to develop a novel 4D convolution network to extract 4D joint temporal-spatial kernels that not only learn spatial information but in addition also capture temporal dynamics.

**Approach:** We apply our novel approach on the ADNI dataset with data augmentations such as circular time shifting to enforce time-invariant results.

**Results:** Experimental results show promising performance in capturing spatial-temporal data in functional MRI compared to 3D models.

**Impact:** The 4D CNN model improves Alzheimer's disease diagnosis for rs-fMRI data, enabling earlier detection and better interventions. Future research could explore task-based fMRI applications and regression tasks, enhancing understanding of cognitive performance and disease progression.

## 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<sup>1,2</sup>. Feature extraction from this neuroimaging data typically involves machine learning algorithms applied to either functional connectivity matrices or subcortical surface maps<sup>3,4</sup>. Contrarily, this study focuses on the 4-dimensional data for classification which is often overlooked due to bigger computational demands. We evaluate three deep learning approaches for handling 4D data. The first approach employs a 3D convolutional neural network (CNN) using the ConvNeXt architecture<sup>5</sup>, treating time samples as input channels. The second approach is a hybrid model combining a 3D CNN with a long short-term memory (LSTM) module<sup>6</sup> to separately capture spatial features and temporal dynamics. The third approach introduces a novel 4D CNN model that performs convolutions using 4D temporal-spatial kernels. While using a 4D CNN is not entirely unprecedented<sup>7,8</sup>, this study represents the first application of such a model in rs-fMRI for diagnosing Alzheimer's disease.

## Methods

We used the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset<sup>9</sup>, selecting 3T rs-fMRI scans (EPI pulse sequence data) with spatial resolution 3.3mm 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 of 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<sup>10</sup> and processing with fmrip<sup>11</sup>. For structural scans, this included N4 bias field correction, skull stripping, and spatial normalization to the MNI152 linear space from TemplateFlow<sup>12</sup>. 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<sup>13</sup>, 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<sup>14</sup> 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<sup>15</sup> 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 Figure 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 Figure 2. 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 Table 1. 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 (Figure 3). From this, it appears that some lower-level features include derivative and averaging information. Secondly, the Grad-CAM++ method<sup>16</sup> was employed to identify important regions used for diagnosis. Figure 4 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.

## Acknowledgements

This work was fully supported by NIH grant R21 AG082204.

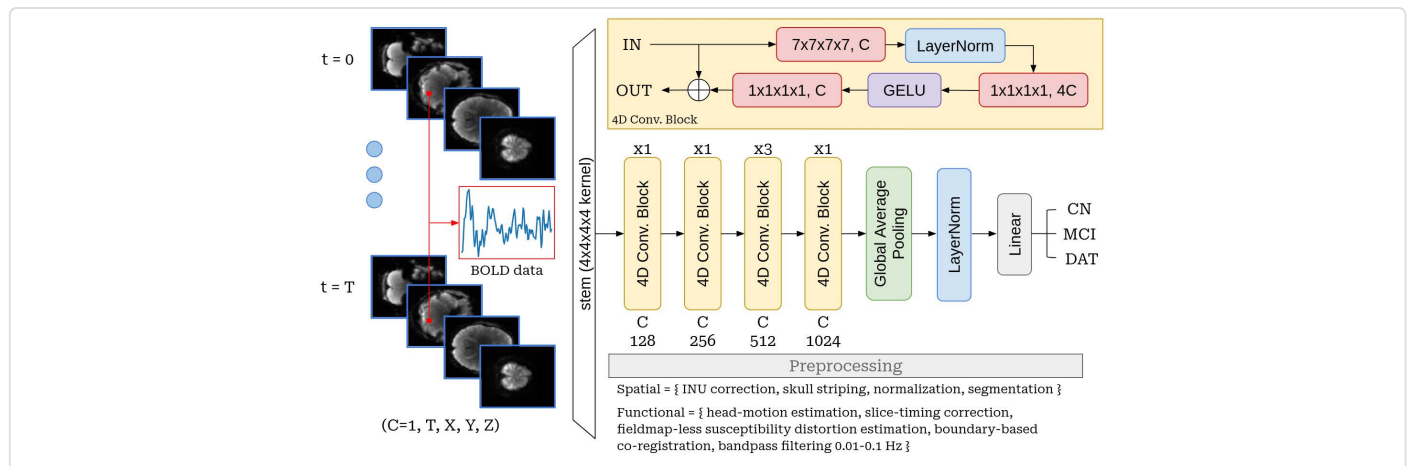
## References

1. Vemuri, P., Jones, D. T., & Jack, C. R. (2012). Resting state functional MRI in Alzheimer's Disease. *Alzheimer's research & therapy*, 4, 1-9.
2. Sperling, R. (2011). The potential of functional MRI as a biomarker in early Alzheimer's disease. *Neurobiology of aging*, 32, S37-S43.
3. Vemuri, P., Jones, D. T., & Jack, C. R. (2012). Resting state functional MRI in Alzheimer's Disease. *Alzheimer's research & therapy*, 4, 1-9.
4. Karker, M. (2022). *Predictive Analysis and Deep Learning of Functional MRI in Alzheimer's Disease* (Doctoral dissertation).
5. Liu, Z., Mao, H., Wu, C. Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). A convnet for the 2020s. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 11976-11986).
6. Yu, Y., Si, X., Hu, C., & Zhang, J. (2019). A review of recurrent neural networks: LSTM cells and

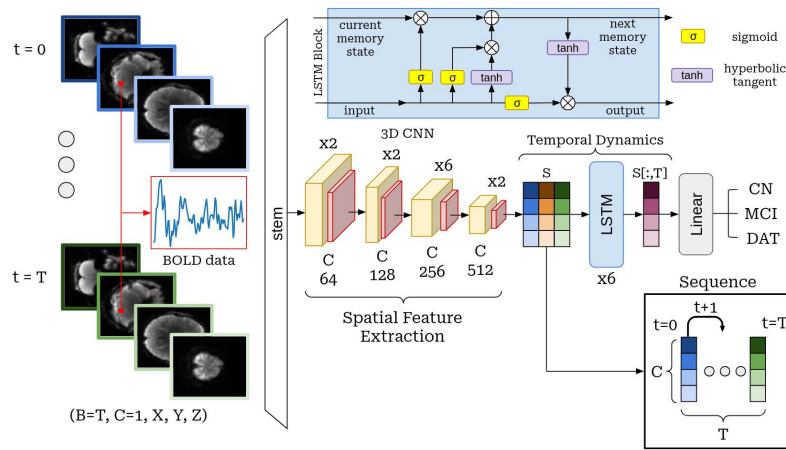
network architectures. *Neural computation*, 31(7), 1235-1270.

7. Myronenko, A., Yang, D., Buch, V., Xu, D., Ihsani, A., Doyle, S., ... & Roth, H. (2020). 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 International Publishing.
8. Bengs, M., Gessert, N., & Schlaefer, A. (2020). 4d spatio-temporal deep learning with 4d fmri data for autism spectrum disorder classification. *arXiv preprint arXiv:2004.10165*.
9. Mueller, S. G., Weiner, M. W., Thal, L. J., Petersen, R. C., Jack, C., Jagust, W., ... & Beckett, L. (2005). The Alzheimer's disease neuroimaging initiative. *Neuroimaging Clinics of North America*, 15(4), 869.
10. Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., ... & Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific data*, 3(1), 1-9.
11. Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., ... & Gorgolewski, K. J. (2019). fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature methods*, 16(1), 111-116.
12. Ciric, R., Thompson, W. H., Lorenz, R., Goncalves, M., MacNicol, E. E., Markiewicz, C. J., ... & Esteban, O. (2022). TemplateFlow: FAIR-sharing of multi-scale, multi-species brain models. *Nature Methods*, 19(12), 1568-1571.
13. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *The Journal of machine Learning research*, 12, 2825-2830.
14. Brett, M., Markiewicz, C. J., Hanke, M., Côté, M.-A., Cipollini, B., Papadopoulos Orfanos, D., McCarthy, P., Jarecka, D., Cheng, C. P., Larson, E., Halchenko, Y. O., Cottaar, M., Ghosh, S., Wassermann, D., Gerhard, S., Lee, G. R., Baratz, Z., Moloney, B., Wang, H.-T., ... freec84. (2024). nipy/nibabel: 5.3.1 (5.3.1). Zenodo. <https://doi.org/10.5281/zenodo.13936989>
15. Kingma, D. P. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
16. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision* (pp. 618-626).

## Figures



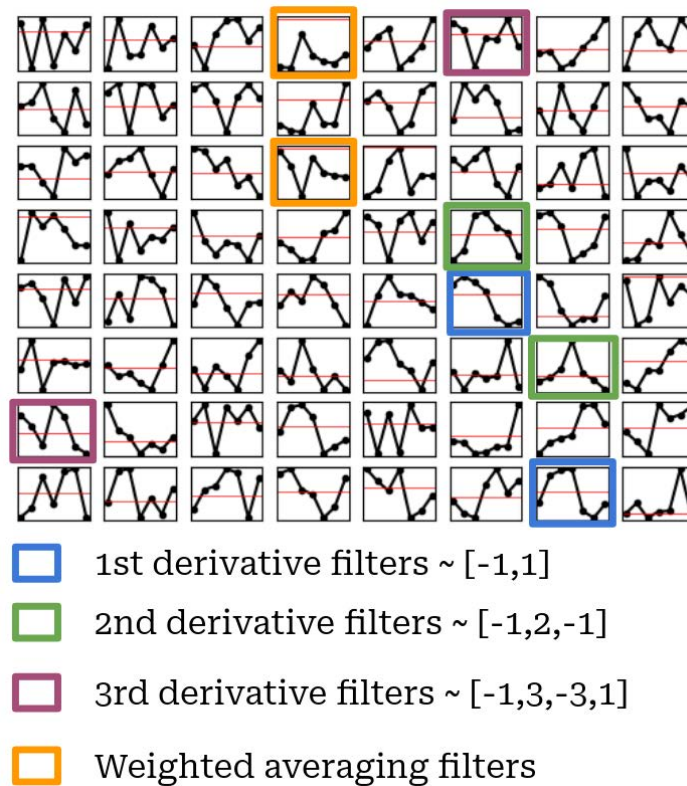
**Figure 1.** Proposed 4D CNN model architecture, consisting of four downsampling stages in a 1-1-3-1 configuration. The final stage outputs 1024 channels that are globally average pooled to yield 1024 features for the entire 4D scan.



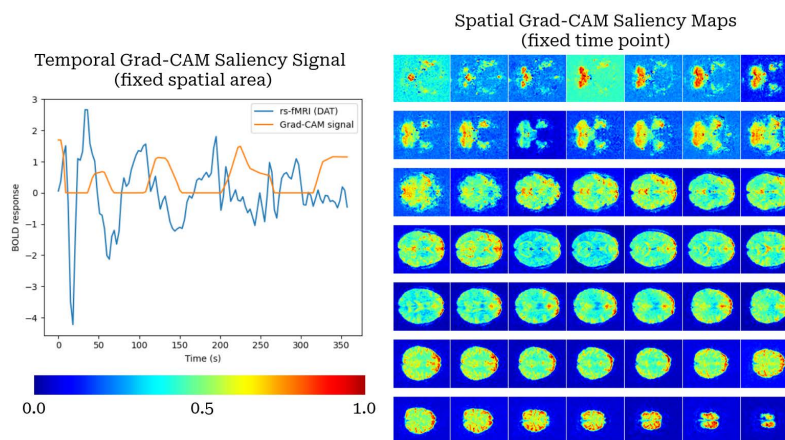
**Figure 2.** Architecture of the hybrid 3D CNN + LSTM model. Each 3D time sample is processed individually by the CNN, and the resulting features are aggregated into matrix  $S$ . The LSTM module captures temporal dynamics between time samples for classification purposes.

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

**Table I.** 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 3.** 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 4.** 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.