

A MULTIMODAL DEEP LEARNING APPROACH FOR AUTOMATED DETECTION AND CHARACTERIZATION OF DISTINCTLY SALIENT FEATURES OF ALZHEIMERS DISEASE

Ishaan Batta, Anees Abrol, Vince Calhoun

Center for Translational Research in Neuroimaging and Data Science (TReNDS):
Georgia State University, Georgia Institute of Technology, and Emory University, Atlanta, USA

ABSTRACT

Neurological disorders generally involve multiple kinds of changes in the functional and structural properties of the brain. In this study, we develop a CNN-based multimodal deep learning pipeline by exploiting both functional and structural neuroimaging features to generate full-brain maps that encode significant differences between patient groups and between modalities in terms of their distinctive contribution towards diagnostic classification of Alzheimer's disease. Through a repeated cross-validation procedure and robust statistical analysis, we show that our approach can be used to encode highly discriminative and abstract information from full-brain data, while also retaining the ability to identify and categorize significantly contributing voxel-level features based on their salient strength in various diagnostic and modality-related contexts. Our results on an Alzheimer's disease classification task show that such approaches can be used for creating more elaborately defined biomarkers for brain disorders.

Index Terms— Deep Learning, Multimodal Fusion, Neuroimaging, Saliency, Alzheimer's Disease

1. INTRODUCTION

Understanding how the brain's structure and function are related to various behavioral and neurodegenerative disorders is crucial for finding solutions to target them. Various neuroimaging studies have shown that utilizing the information in high-dimensional measures from the brain can be useful toward this goal. In the past decade, various learning approaches have been developed for discriminating various brain disorders from neuroimaging data. Additionally, there has been a lot of interest in studying whether the information encoded by these approaches is relevant to understand the corresponding disorder.

While standard machine learning approaches have been employed toward such interpretations, they are either more suited to higher-level features that are generalized at the level of brain regions or perform sub-optimally for meaningful performance on high-dimensional data [1]. Moreover, with rising evidence regarding both structural and functional properties of the brain being affected in various brain disorders,

it is increasingly important to have frameworks that can fuse both structural and functional neuroimaging features [2]. For this, one of the challenges is to have frameworks that extract meaningful low-dimensional representations from the increased multimodal scale of the feature dimensions, as well as retain meaningful patterns in sufficient detail to interpret the diverse and subtle changes in the brain successfully.

Deep learning approaches have been shown to successfully encode discriminative features leading to a more robust predictive analysis under various scenarios [1]. With deep learning methods based on convolutional neural networks (CNNs), the spatial information from a neuroimaging scan can be exploited all the way from the basic voxel level up to higher level association. Furthermore, using various saliency approaches [3], the learned highly discriminative information can be projected back from highly abstract and low-dimensional space back onto the level of voxels. Such analyses have opened up a high potential for creating detailed interpretations using neuroimaging data for various unanswered problems in neuroscience. Given appropriate statistical measures are developed for the same, CNN-based deep learning methods can aid in more robust and detailed biomarker discovery for brain disorders.

With the same motivation, we use a CNN-based deep learning approach to study the underlying differences in the functional as well as structural features responsible for distinguishing Alzheimer's disease (AD). We first train a multimodal CNN architecture created by modifying AlexNet [4] for AD classification by synthesizing various functional and structural measures from fMRI and MRI data, respectively. This is followed by the computation of full-brain saliency maps using a gradient back-propagation approach [3]. Subsequently, through a repeated cross-validation procedure, we analyze the voxel-level differences in the full-brain saliency for (a) control subjects vs. subjects with Alzheimer's disease and (b) functional and structural modalities. Through a robust statistical testing procedure, we visualize brain maps representing significant diagnostic and modality-wise differences in the saliency of each spatial location in the brain. Lastly, we compute measures to summarize the strength of these differences at the level of brain regions. Our analysis reveals that regions well known to be biologically associated with AD can

Corresponding author: Ishaan Batta (ibatta@gatech.edu)

be successfully uncovered using DL approaches. Moreover, our analysis reveals that regions can be attributed differently under different diagnostic or modality-related contexts based on their preferential saliency toward distinguishing AD.

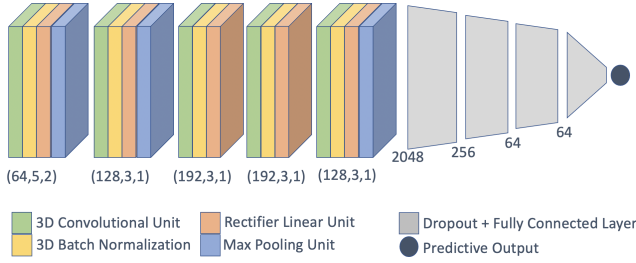


Fig. 1: The multimodal CNN architecture used for the analysis. Voxel-level brain maps for structural and functional features were used as multi-channel input to the architecture with shared parameters in the convolutional layers, followed by fully connected layers for classification. This was followed by the computation of subject-specific saliency maps and statistical analysis to compare differences between diagnostic groups and modalities.

2. METHODS

2.1. Dataset and Pre-Processing

Functional and structural MRI data from the ADNI dataset (adni.loni.usc.edu) were used for only the first visit scans for 506 subjects (214/292 M/F, aged 71.61 ± 6.76) with 415 controls (CN) and 91 subjects with Alzheimer’s Disease (AD). Preprocessing was done using a standard SPM12 pipeline as in previous studies [5]. Functional measures computed for the analysis from fMRI data included amplitude of low-frequency fluctuations (ALFF), regional homogeneity (ReHo) using Kendall’s coefficient of concordance, weighted degree centrality (DCw) [6], and the average of the fMRI time-series (tsavg), while the structural measures included low-resolution T1-weighted images (IT1). For feeding into the multimodal deep learning architecture, all maps were warped to the standard MNI space, resampled to $(3\text{mm})^3$ isotropic voxels with $53 \times 63 \times 52$ voxels in each map, followed by Gaussian smoothing (FWHM = 6mm).

2.2. Multimodal Deep CNN Classifier

The preprocessed feature maps were used as input to a multi-channel variant of the AlexNet architecture [4] shown in 1, which is known to successfully encode predictive features from neuroimaging data [1]. For all four fMRI and the sMRI feature maps, three cases of possible ways to use them in the architecture were evaluated for comparison purposes. These include (a) unimodal, with only a single channel using only one of the feature maps, (b) 2-way multimodal, with IT1 and one of the fMRI feature maps being used as input. For the multimodal cases, shared parameters were used in convolutional layers for to ensure collective feature extraction in the multi-channel architecture. The analysis was done on tuned parameters (batch-size = 32, learning rate = 0.001), and was

repeated for 10 repetitions using a random sub-sampling procedure by dividing the data into stratified training, validation, and test (in proportion 3:1:1) for each repetition.

2.3. Saliency Analysis and Comparison

Saliency maps were computed for each test subject using guided back-propagation [3] on the learned architecture. It should be noted that in the case of multimodal architectures, the saliency maps are computed for each of the input feature maps. To compare the saliency between CN and AD groups, the maps from subjects in each of the classes were standardized by taking a z-score across voxels for each subject and tested for statistically significant differences using a two-sample t-test followed by false discovery rate (FDR) correction using the Benjamini and Hochberg procedure ($\alpha = 0.01$). The voxel-wise corrected p -values < 0.01 were used as a measure for indicating a statistically significant difference between the saliency of the two groups. For comparing which brain areas have differences in functional vs structural saliency, a similar procedure was performed between the saliency maps of the two features used within the same 2-way multimodal learning paradigm.

Unimodal acc ($\mu \pm \sigma$)		Multimodal acc ($\mu \pm \sigma$)	
IT1	.90 $\pm$.02		
tsavg	.82 $\pm$.02	tsavg-IT1	.88 $\pm$.04
ALFF	.83 $\pm$.01	ALFF-IT1	.89 $\pm$.02
DCw	.82 $\pm$.007	DCw-IT1	.88 $\pm$.03
ReHo	.82 $\pm$.006	ReHo-IT1	.89 $\pm$.03

Table 1: Comparison of performance based on accuracy scores for the cases of unimodal and multimodal architectures for 10 repetitions of analysis with a random-subsampling procedure to create separate folds for training, validation, and held-out test data. The functional measures used were tsavg, ALFF, DCw, and ReHo, while the T1-weighted maps (IT1) registered to fMRI domain were used as a structural measure. For functional measures, multimodal setting yields better performance than the unimodal one.

3. RESULTS

3.1. Performance Comparison

Table 1 shows the test accuracy on 10 repetitions of the experiment using a repeated stratified sub-sampling procedure for both unimodal as well as multimodal combinations of the involved features. It can be noted that while the structural features perform better than the functional features in a unimodal setting, the performance in multimodal settings is not significantly different ($p > .05$ for all cases). Another interesting aspect of the above results is that compared to other functional measures, tsavg performs equally well, indicating that the average fMRI activation at each voxel also inherently encodes sufficient discriminatory information.

3.2. Distinctive Saliency

Saliency maps were computed and statistical comparison was done for groups (CN vs AD) and modalities (functional vs

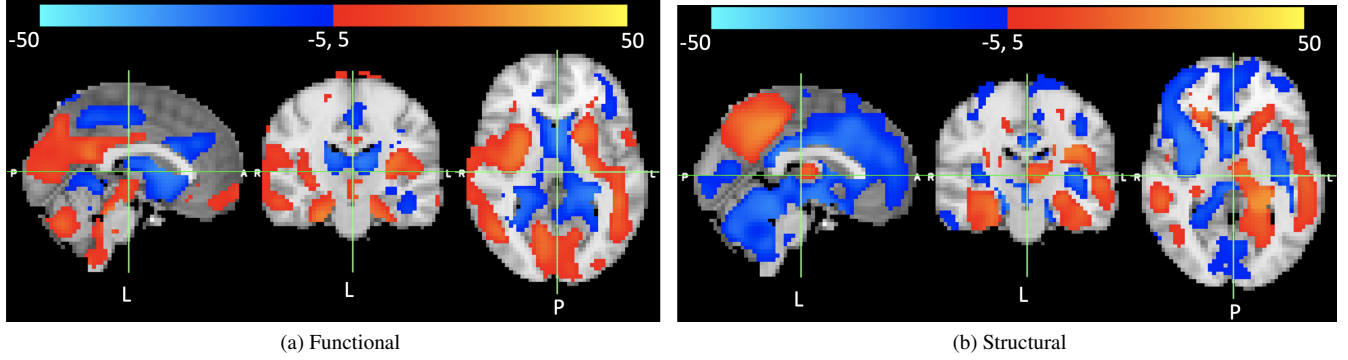


Fig. 2: Full-brain maps showing FDR-corrected $-sign(t) \log(p)$ values for two-sample t-test for difference between the mean saliency of CN vs. AD subjects at each voxel for standardized saliency maps. Results are shown for the case of multimodal CNN architecture described in subsection 2.2 using functional (ALFF) and structural (IT1) feature maps as inputs, resulting in two maps per subject which were tested separately for CN vs AD differences. A positive value at a voxel implies significantly stronger saliency in CN than in AD group. Maps for other functional features are omitted due to space constraints, but top brain regions from those cases can be found in Table 2.

ALFF-IT1		DCw-IT1		ReHo-IT1		tsavg-IT1	
ALFF	IT1	DCw	IT1	ReHo	IT1	tsavg	IT1
LThal (-22.1)	LAmyg (26.53)	OccPole (18.82)	LAmyg (21.83)	RAmyg (-19.6)	LAmyg (23.85)	LAmyg (-17.36)	LAmyg (23.48)
RThal (-18.45)	RAmyg (23.19)	LThal (-16.79)	RAmyg (20.13)	iCalc (19.05)	RAmyg (21.65)	CB11-4 (15.44)	LHipp (19.91)
RCaud (-17.82)	CBv7-10 (-21.75)	CBICr (16.56)	TfusiA (19.93)	LAmyg (-18.66)	LHipp (20.84)	RNA (-14.1)	RAmyg (19.53)
RLatV (-17.62)	LHipp (21.21)	LOCi (16.5)	CBvCr (-17.95)	SupCalc (18.41)	CBv7-10 (-17.71)	RPutA (13.23)	TfusiA (19.2)
LCaud (-16.83)	TfusiA (20)	RPall (-15.91)	SFG (-17.72)	LThal (-16.43)	CBv7-10 (-17.07)	TfusiA (-13.22)	CingGa (-17.61)
RPutA (15.91)	LBrSt (-19.96)	SpmGa (15.59)	OccPole (-17.72)	CBICr (16.35)	PrHippGa (16.48)	CB15-6 (13.03)	PrHippGa (17.27)
LLatV (-15.7)	CingGa (-17.83)	RAmyg (-15.23)	FP (-17.04)	CBv7-10 (15.8)	FP (-16.47)	CBv5-6 (12.03)	CBv7-10 (-16.95)
CB17-10 (15.09)	PrHippGa (17.83)	CBrCr (14.51)	ParaC (-16.81)	CBv7-10 (15.44)	IFGpt (-16.21)	MTGa (-11.64)	ParaC (-15.66)
LNA (-14.79)	RNA (-17.16)	FP (13.87)	PrHippGa (16.66)	ITGa (15.43)	CingGa (-16.2)	PrHippGp (-11.59)	LBrSt (-15.1)
LPall (14.7)	CBv7-10 (-17.01)	ITGa (13.74)	SMC (-16.46)	RThal (-15.14)	LBrSt (-15.92)	TP (-11.35)	RPall (-14.61)

Table 2: Top 10 ROIs with the highest strength of CN vs. AD saliency differences. For all combinations of multimodal (functional and structural) input features, CN vs. AD comparison results were computed separately for each modality within a given combination. For a given ROI r , the saliency strength $\bar{s}_r$ is defined as the mean of $|-sign(t) \log(p)|$ values for all the voxels $v \in r$, where the t, p values are results of FDR-corrected two-sample t-test between the mean saliency for CN vs AD subjects. While $\bar{s}_r$ is always positive, the sign of the $\bar{s}_r$ values in parentheses for each ROI indicates the sign t-statistic for the majority of the voxels in the ROI that survived FDR-correction. Thus, a positive sign indicates a higher saliency in the CN group than in AD for most of the significant voxels in that ROI. Full names of ROIs can be found at this link

structural) as described in 2.3. To better visualize the statistical differences, $-sign(t) \log(p)$ values from the two-sample t-test after FDR correction were plotted as a brain map using the FSL toolbox [7]. For ranking various brain regions of interest (ROIs) based on these voxel-level comparison maps, the mean strength $\bar{s}_r = \sum_{v \in r} |-sign(t) \log(p)|$ across all voxels in a given ROI ($v \in r$) was used as a metric. While the metric $\bar{s}_r$ cannot be negative by definition, a positive or negative sign was ascribed in the visualizations based on the $sign(t)$ value for the majority of the voxels in the ROI that survived FDR-correction (Figure 2, Table 2). ROIs were defined by combining non-overlapping regions in Harvard-Oxford cortical, sub-cortical, and cerebellar atlases available in FSL [7].

Figure 2 shows the maps for group differences (denoted by $-sign(t) \log(p)$ values) in the mean saliency between CN and AD groups in a 2-way multimodal classifier. The top 10

ROIs based on the value of $\bar{s}_r$ are shown in Table 2. Secondly, to visualize how the functional and structural modalities differ in terms of the most salient brain areas under the same 2-way multimodal training of the model, Figure 3 summarizes the results from FDR-corrected two-sample t-test between mean saliency of IT1 and ALFF features.

3.3. Relevant Brain Areas

It can be noted that the set of regions with the most significant CN vs AD differences have a mix of unique as well as common ROIs between functional and structural saliency cases. Similarly, multiple brain areas show significant differences in saliencies when using functional vs structural features.

Brain regions, including the Amygdala, Hippocampus, para-Hippocampal Gyrus, Fusiform gyrus, Occipital Pole, and Cerebellar areas, show the strongest differences in the saliency of CN vs. AD groups. Additionally, in terms of the

CN vs. AD differences in the saliency of the functional measures, it can be noted that Thalamus, Caudate, and Putamen feature in almost all cases of fMRI-based measures (Table 2). These results are in line with the previous findings about regions that are known to be disrupted in Alzheimer’s disease [8]. Moreover, the comparison between ALFF and IT1 saliencies (Figure 3) reveals that ALFF has mainly subcortical areas with the strongest saliency difference, while both temporal and subcortical regions feature in the case of IT1.

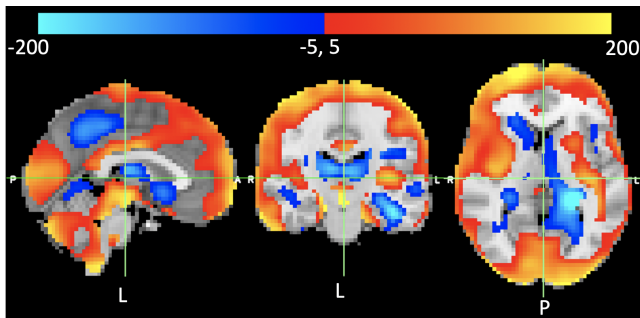


Fig. 3: Full-brain maps for FDR-corrected $-sign(t) \log(p)$ values for two-sample t-test between mean voxel-level saliency of functional (ALFF) and structural (IT1) features in a multimodal architecture trained as in 2.2. The top 10 ROIs with the highest mean strength as defined in 3.2 were OccPole (148.29), LAmyg (-131.96), LHipp (-124.11), CBICr (104.55), TfusiP (-103.69), PoCG (100.93), CBI7-10 (99.42), FP (99.21), SpmGa (96.92), SFG (94.0). For both the brain maps and ROIs listed, a positive value implies a significantly higher strength of functional saliency than structural for the given voxel/ROI. Full names of ROIs can be found at this link

4. CONCLUSION

In this work, we propose a way to identify the brain regions that show inter-group as well as inter-modality differences in the saliency toward predicting a given brain disorder. By analyzing the saliency results with appropriate statistical tests, we show that deep learning methods can be employed to have more informed as well as diverse interpretations of salient high-dimensional neuroimaging features. By combining multiple features from functional as well as structural domains into a single framework, both associations and differences can be uncovered between the functional and structural aspects of the brain under both health and disease.

In conclusion, our results indicate that the discriminative associations encoded by deep learning models for diagnostic classification are in line with the neurobiological findings about the brain areas known to be affected both structurally as well functionally in Alzheimer’s disease. While the scope of this paper had to be limited to specific methods for classification, saliency, and comparison, future work could involve a more detailed study to elaborate the perspective along multiple dimensions of analysis.

5. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted retrospectively using human subject data made available in open access by ADNI

(LONI). Ethical approval was not required and approval documents for ADNI data can be accessed via (adni.loni.usc.edu).

6. ACKNOWLEDGMENTS

The work reported in this manuscript was supported by grants from the National Science Foundation (NSF 2112455) and the National Institutes of Health (NIH RF1AG063153).

7. REFERENCES

- [1] Anees Abrol, Zening Fu, Mustafa Salman, Rogers Silva, Yuhui Du, Sergey Plis, and Vince Calhoun, “Deep learning encodes robust discriminative neuroimaging representations to outperform standard machine learning,” *Nature communications*, vol. 12, no. 1, pp. 1–17, 2021.
- [2] Vince D Calhoun and Jing Sui, “Multimodal fusion of brain imaging data: a key to finding the missing link (s) in complex mental illness,” *Biological psychiatry: cognitive neuroscience and neuroimaging*, vol. 1, no. 3, pp. 230–244, 2016.
- [3] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman, “Deep inside convolutional networks: Visualising image classification models and saliency maps,” *arXiv preprint arXiv:1312.6034*, 2013.
- [4] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, “Imagenet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.
- [5] Anees Abrol, Manish Bhattarai, Alex Fedorov, Yuhui Du, Sergey Plis, Vince Calhoun, Alzheimer’s Disease Neuroimaging Initiative, et al., “Deep residual learning for neuroimaging: an application to predict progression to alzheimer’s disease,” *Journal of neuroscience methods*, vol. 339, pp. 108701, 2020.
- [6] Xi-Nian Zuo, Ross Ehmke, Maarten Mennes, Davide Imperati, F Xavier Castellanos, Olaf Sporns, and Michael P Milham, “Network centrality in the human functional connectome,” *Cerebral cortex*, vol. 22, no. 8, pp. 1862–1875, 2012.
- [7] Stephen M Smith, Mark Jenkinson, Mark W Woolrich, Christian F Beckmann, Timothy EJ Behrens, Heidi Johansen-Berg, Peter R Bannister, Marilena De Luca, Ivana Drobnjak, David E Flitney, et al., “Advances in functional and structural mr image analysis and implementation as fsl,” *Neuroimage*, vol. 23, pp. S208–S219, 2004.
- [8] Zhengjia Dai and Yong He, “Disrupted structural and functional brain connectomes in mild cognitive impairment and alzheimer’s disease,” *Neuroscience Bulletin*, vol. 30, no. 2, pp. 217–232, 2014.