2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)
April 4-7, 2018, Washington, D.C., USA

DEEP FMRI: AN END-TO-END DEEP NETWORK FOR CLASSIFICATION OF FMRI DATA

Atif Riaz *Muhammad Asad* *S M Masudur Rahman Al Arif*
Eduardo Alonso *Danai Dima* *Philip Corr* *Greg Slabaugh*
City, University of London, UK

ABSTRACT

With recent advancements in machine learning, the research community has made tremendous advances towards the classification of neurological disorders from time-series functional MRI signals. However, existing classification techniques rely on hand-crafted features and classical machine learning models. In this paper, we propose an end-to-end model that utilizes the representation learning capability of deep learning to classify a neurological disorder from fMRI data. The proposed DeepFMRI model is comprised of three networks, namely (1) a feature extractor, (2) a similarity network, and (3) a classification network. The model takes fMRI raw time-series signals as input and outputs the predicted labels; and is trained end-to-end using back-propagation. Experimental results on the publicly available ADHD-200 dataset demonstrate that this innovative model outperforms previous state-of-the-art.

Index Terms— Deep learning, end-to-end model, fMRI classification

1. INTRODUCTION

In recent years, functional magnetic resonance imaging (fMRI) has emerged as a popular neuroimaging modality for classification of neurological disorders. Specifically, resting state fMRI has emerged as a powerful tool to study the functional organization of the brain. Many studies [1–3] have shown promising outcomes in the classification of brain disorders like attention deficit hyperactivity disorder (ADHD), schizophrenia and Alzheimer's disease by studying brain functional networks in resting state fMRI. fMRI data can be viewed as a 4D tensor such that the 3D volume of the brain is divided into small voxels or regions and the activity of each region is recorded for a certain duration. Two brain regions that show synchronous functional activity are assumed to be functionally connected. Functional connectivity is viewed as a pair-wise connectivity measurement that describes the strength of temporal coherence between the brain regions. A number of recent studies have shown functional connectivity as an important biomarker for discrimination of different brain disorders like ADHD [1], schizophrenia [3] and many more.

ADHD is one of the most common neuro-developmental and mental disorders affecting 5-10% of school going children [2], contributing to lifetime impairment [4], poor quality of life [5] and long-term burden on affected families [4, 5]. Like many other neurological disorders, the underlying mechanism of ADHD is still unknown [2]. As there is no single confirmed diagnostic method available for ADHD, diagnosis is dependent upon observations conducted by medical practitioners or parents, typically over a period of months.

Several techniques have applied hand-crafted features for classification of ADHD from fMRI data, such as correlation [3], clustering [1] and graph [2] measures of functional connectivity. Discriminant features are selected and presented to a classical machine learning classifier for final prediction. However, in the machine learning literature, deep learning has proved to be a powerful paradigm to simultaneously learn discriminant features and a classifier [6].

End-to-end deep learning networks have been shown to outperform classical machine learning models in a number of domains like image classification, image segmentation and object recognition [6]. Generally speaking, an end-to-end trainable network refers to a single learning system where the predicted label of a machine learning process is predicted directly from the raw input, with all weights learned through back-propagation. Recently, a deep learning method named FCNet [7] has been proposed for classification of ADHD from fMRI data. The method uses a convolutional neural network (CNN) to predict functional connectivity of brain regions. However, after predicting functional connectivity using deep learning, the method uses classical machine learning methods to extract discriminant features and an SVM classifier to predict classification labels.

To our knowledge, this paper presents the first end-to-end deep learning model for classification of a neurological disorder from fMRI data. Particularly, we are interested to see if a deep learning model can be designed for the classification of a neurological disorder and if it is able to outperform classical machine learning models.

The architecture of DeepFMRI is illustrated in Fig. 1. The proposed model is inspired by the FCNet [7] and uses a pre-trained FCNet for some of the initial layers to extract features from the raw fMRI time-series signals. Unlike FCNet, these

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan X Pascal GPU used for this research.

978-1-5386-3636-7/18/$31.00 ©2018 IEEE

layers are fine-tuned during training to learn the features of individual brain regions. During training, the end-to-end model learns weights to distinguish between the healthy control and ADHD subjects. Once the model is trained, unseen data is provided and the model predicts the classification label.

The contributions of this paper include: 1) a novel deep learning end-to-end model that simultaneously learns the most discriminant features and a classifier to classify fMRI data, and 2) an improved classification accuracy over the state-of-the-art on the ADHD-200 dataset.

2. METHODS

2.1. Data and preprocessing

The resting state fMRI data used in this work is provided by the ADHD-200 consortium [8]. The dataset is acquired by different imaging sites and is comprised of resting state fMRI data, MRI data, as well as phenotypic information. The consortium has provided a training dataset, and an independent testing dataset separately for each individual imaging site. In this work, we have used resting state fMRI data from three sites: NeuroImage (NI), New York University Medical Center (NYU) and Peking University (Peking). All imaging sites in the consortium have a different number of subjects. Additionally, imaging sites have different scan parameters and equipment, which makes the dataset complex and diverse for building any machine learning model. This data has been preprocessed as part of the connectome project[1]. The preprocessing involved different steps where the brain is segmented into 90 regions using the automated anatomical labeling (AAL) atlas [9]. A more detailed description of the data and preprocessing steps appears on the connectome website. Each segmented region is represented by a time-series signal that captures the level of blood oxygenation (BOLD signal). These 90 time series signals are the input to DeepFRMI.

2.2. End-to-end model

In this paper, we propose an end-to-end deep learning model for classification of ADHD that takes fMRI signals as input and predicts a label (1 for ADHD subject and 0 for healthy control) as output. The proposed work is motivated by a recently published method called FCNet [7]. FCNet is used to extract functional connectivity from fMRI time-series signals, but it suffers from the following drawbacks: i) it is not an end-to-end model, and ii) it relies on classical machine learning methods like feature selection using elastic net and a support vector machine for classification. Our DeepFMRI model architecture can be divided into three modules: i) feature extractor, ii) similarity measure, and iii) classification network. DeepFMRI employs multiple FCNets (with shared parameters) for calculating functional connectivity for any pair of brain regions. The FCNets are fine-tuned and combined with a classification network to provide a fully end-to-end model that can be trained using back-propagation. We describe the details of each individual network below.

[1] www.preprocessed-connectomes-project.org/adhd200/

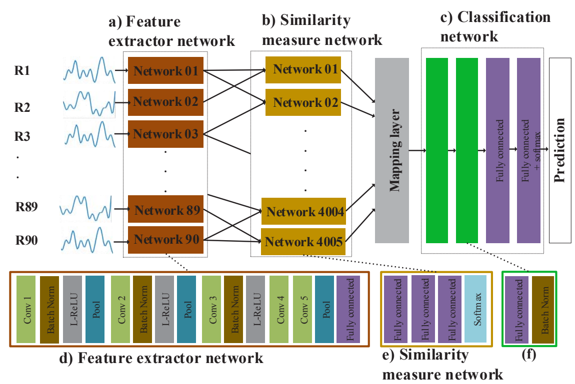

Fig. 1. The architecture of the proposed end-to-end model. a) represents a set of 90 feature extractor networks where each network is applied to each individual region R. All networks share same parameter set. b) represents a set of 4005 similarity measure networks. Each network's input contains abstracted features of two brain regions. All networks share the same parameter set. c) is the classification network comprising of fully connected layers and a softmax layer. d) represents the details of layers in feature extractor network. Similarly, e) represents layer architecture of similarity measure network, and f) represents the detail of layers of the individual block in the classification network (the two blocks in the classification network do not share parameters).

2.2.1. *The feature extractor*

This convolutional neural network extracts features from *individual* brain region time-series signals and is comprised of multiple layers that are common in CNN models to learn abstract representations of the individual time-series signal. The network is designed to accept time-series signals of length 172 and is comprised of multiple layers (presented in Fig. 1c). For all convolutional layers, a kernel size of 3 is used and the number of filters are 32, 64, 96, 64, 64 for layers $Conv1$, $Conv2$, $Conv3$, $Conv4$, $Conv5$, respectively. All pooling layers pool temporally with pool length of 2. The last fully connected layer in the network has 32 nodes.

2.2.2. *The similarity measure network*

This Siamese-inspired neural network determines the similarity between *pairs* of extracted features from two brain regions. The output of this network describes the degree of functional connectivity between the two regions. The input to this network is the abstracted features extracted for the two regions through the feature extractor network. The similarity measure network is comprised of three fully connected layers, where the last layer is connected to a softmax layer with dense connections. These layers are presented in Fig 1e.

Similarity measures are fed to a mapping layer with following operation:

$$M(i) = w_1 v_1^i + w_2 v_2^i, \tag{1}$$

where v_1^i and v_2^i are the outputs of i^{th} similarity measure network, w_1 and w_2 are the weights such that $w_1 + w_2 = 1$, here we use $w_1 = 1$ and $w_2 = 0$. Instead of initializing weights of the feature extractor network and similarity measure network randomly, we use weights of a pre-trained FCNet [7].

2.2.3. *Classification network*

This neural network produces the final classification results. The input to this network is the output of the mapping layer features (M) representing functional connectivity. The network is comprised of multiple layers where the last layer is connected to a softmax classifier with fully connected layers. The network produces the final prediction. Next, we describe architectural considerations and training of DeepFMRI.

2.2.4. *Shared parameters architecture*

The architecture of the feature extractor network and similarity measure network is the same as FCNet. However, the FCNet architecture cannot be applied directly to construct an end-to-end network as it is designed to work on only two brain regions. In DeepFMRI, the same feature extraction steps are applied to individual brain regions, and all pairs of brain regions are passed through the same similarity measure network. This is realized by employing n_f feature extractor networks and n_s similarity measure networks. Each feature extractor network is applied to an individual brain region ($n_f =$ 90), converting individual time-series data into an abstract representation. All the feature extractor networks share the same parameters and updates are applied to these shared parameters during training. The similarity measure network is applied to all combinations of pairs of brain regions, so $n_s = 4005$ ($n_f \times (n_f - 1)/2$). All the similarity measure networks are implemented with the constraint that the networks share the same parameters and updates are applied to these shared parameters. The approach is similar to a Siamese network [10], however, typically Siamese networks are designed to work on image pairs.

3. EXPERIMENTS AND RESULTS

The proposed DeepFMRI model is evaluated on the ADHD-200 dataset. The dataset was contributed by different imaging sites. Each imaging site provided separate training and testing dataset. For evaluation of our method on individual site, we train our end-to-end model on the training dataset of each imaging site and test on the corresponding test dataset of that individual site. There are four categories of subjects

Table 1. Results from the proposed end-to-end network showing classification accuracy, specificity and sensitivity.

	Classification accuracy	Specificity	Sensitivity
NYU	73.1%	91.6%	65.5%
NI	67.9%	71.4%	63.6%
Peking	62.7%	79.1%	48.1%

in the dataset: healthy control, ADHD combined, ADHD hyperactive-impulsive and ADHD inattentive. Here, we combine all ADHD types in one category to investigate classification between healthy control and ADHD.

The network is trained end-to-end. For initialization of the feature extractor and similarity measure networks, we use weights from a pretrained FCNet in our work [7], and these weights are updated through fine-tuning. The end-to-end model is trained with the following loss:

$$L = -\frac{1}{n}\sum_{1}^{n}[y_i log(\hat{y_i}) + (1 - y_i) log(1 - \hat{y_i})], \tag{2}$$

where n is the number of training samples, y_i is the ground truth label of subject (1 for ADHD subject and 0 for healthy control) and $\hat{y_i}$ is the prediction by the proposed network.

As the feature extraction and similarity measure networks are initialized with a pre-trained FCNet, we employ different learning rates for i) feature extraction and similarity measure networks (10^{5}), and ii) the classification network (10^{4}). We evaluate DeepFMRI with data from three imaging sites (NYU, NI and Peking). The number of training subjects in each sites are $226, 48$ and 85 respectively. The results are presented in Table 1. The results show that NYU yields the best result. Table 2 compares our results with the state-of-the-art. The results show that our method outperforms the average accuracy results of competition teams (data from the competition website), highest accuracy for any individual site (from [11]), correlation-based functional connectivity results and clustering based results. Our method also performs well in comparison with the state-of-the-art FCNet method [7]. For correlation results, functional connectivity is calculated through correlation, followed by the elastic net as feature selection and an SVM as the classifier.

Finally, in order to study the differences between the healthy control group and the ADHD group, we visualize their respective functional connectivity difference patterns using the NYU dataset and present the results in Fig. 2. The results show that in ADHD, the frontal lobe functional connectivity is altered the most in ADHD in this dataset.

4. CONCLUSIONS

In this paper, we have proposed an innovative end-to-end convolutional neural network-based deep learning model for

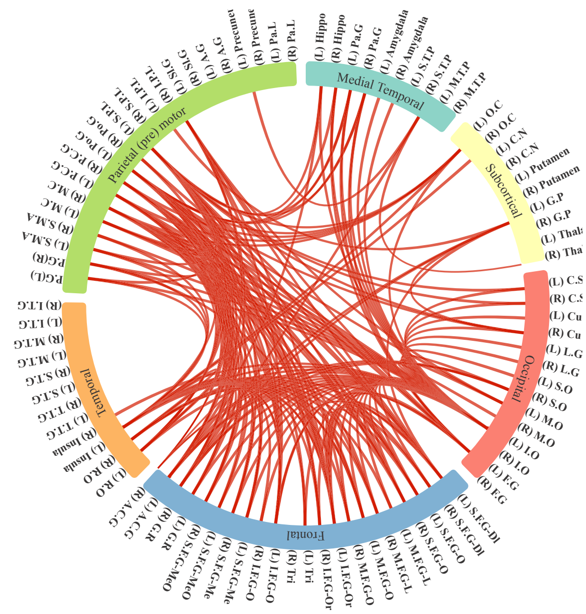

Fig. 2. Visualization of connectivity difference between healthy control and ADHD group for NYU dataset. For the sake of the clarity, only the top 200 connections (based on the connectivity strength) are presented.

Table 2. Comparison of the proposed DeepFMRI method with the average results of competition teams, the highest accuracy achieved for the individual site, correlation based functional connectivity method, clustering based results [1] and state-of-the-art FCNet method [7]. The highest accuracy for NI was not quoted by [11].

	NI	Peking	NYU
Average accuracy [8]	56.9%	51.0%	35.1%
Highest accuracy [11]	–	58%	56%
Clustering method [1]	44%	58.8%	24.3%
Correlation	52.0%	52.9%	56.1%
FCNet [7]	60.0%	**62.7%**	58.5%
DeepFMRI	**67.9%**	**62.7%**	**73.1%**

classification of ADHD from fMRI data. The proposed model takes raw time-series signals of fMRI as input and learns to predict the classification label directly from the raw input values. We were interested to see if the classification task in fMRI can be solved by an end-to-end network. According to our literature study, it is the first attempt to apply an end-to-end network for classification of a neurological disorder. The proposed end-to-end network contains several layers common in deep learning literature. Experimental results on the ADHD-200 dataset demonstrate that utilizing such model outperforms the current state-of-the-art.

5. REFERENCES

[1] Atif Riaz, Eduardo Alonso, and Greg Slabaugh, "Phenotypic Integrated Framework for Classification of ADHD using fMRI," in *International Conference Image Analysis and Recognition*. Springer, 2016, pp. 217–225.

[2] Soumyabrata Dey, A Ravishankar Rao, and Mubarak Shah, "Attributed graph distance measure for automatic detection of attention deficit hyperactive disordered subjects," *Frontiers in Neural Circuits*, vol. 8, 2014.

[3] Junghoe Kim, Vince D Calhoun, Eunsoo Shim, and Jong-Hwan Lee, "Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: Evidence from whole-brain resting-state functional connectivity patterns of schizophrenia," *NeuroImage*, vol. 124, pp. 127–146, 2016.

[4] Valerie A Harpin, "The effect of adhd on the life of an individual, their family, and community from preschool to adult life," *Archives of disease in childhood*, vol. 90, no. suppl 1, pp. i2–i7, 2005.

[5] Joel T Nigg, "Attention-deficit/hyperactivity disorder and adverse health outcomes," *Clinical psychology review*, vol. 33, no. 2, pp. 215–228, 2013.

[6] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[7] Atif Riaz and et al., "FCNet: A Convolutional Neural Network for Calculating Functional Connectivity from functional MRI," in *International Workshop on Connectomics in Neuroimaging*. Springer, 2017, pp. 70–78.

[8] ADHD-200, ," `http://fcon_1000.projects.nitrc.org/indi/adhd200/`.

[9] Tzourio-Mazoyer and et al., "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain," *NeuroImage*, vol. 15, no. 1, pp. 273–289, 2002.

[10] Jane Bromley, Isabelle Guyon, Yann LeCun, Eduard Säckinger, and Roopak Shah, "Signature Verification using a "Siamese" Time Delay Neural Network," in *Advances in Neural Information Processing Systems*, 1994, pp. 737–744.

[11] Marta Nuñez-Garcia, Sonja Simpraga, Maria Angeles Jurado, Maite Garolera, Roser Pueyo, and Laura Igual, "FADR: Functional-Anatomical Discriminative Regions for Rest fMRI Characterization," in *International Workshop on Machine Learning in Medical Imaging*. Springer, 2015, pp. 61–68.