ARTICLE IN PRESS

Neurocomputing xxx (xxxx) xxx

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Deep learning for brain disorder diagnosis based on fMRI images ☆

Wutao Yin [a], Longhai Li [b], Fang-Xiang Wu [c,*]

[a] Division of Biomedical Engineering, 57 Campus Dr., Saskatoon, SK S7N 5A9, Canada
[b] Department of Mathematics and Statistics, University of Saskatchewan, 106 Wiggins Road, MCLN 219 Saskatoon, Saskatchewan S7N 5E6, Canada
[c] Department of Mechanical Engineering, Division of Biomedical Engineering, and Department of Computer Science, 57 Campus Dr., Saskatoon, SK S7N 5A9, Canada

ARTICLE INFO

Article history:
Received 13 March 2020
Revised 17 May 2020
Accepted 22 May 2020
Available online xxxx

Keywords:
Artificial intelligence
Deep learning
Machine learning
Functional magnetic resonance imaging
Brain disorders
Diagnosis

ABSTRACT

In modern neuroscience and clinical study, neuroscientists and clinicians often use non-invasive imaging techniques to validate theories and computational models, observe brain activities and diagnose brain disorders. The functional Magnetic Resonance Imaging (fMRI) is one of the commonly-used imaging modalities that can be used to understand human brain mechanisms as well as the diagnosis and treatment of brain disorders. The advances in artificial intelligence and the emergence of deep learning techniques have shown promising results to better interpret fMRI data. Deep learning techniques have rapidly become the state of the art for analyzing fMRI data sets and resulted in performance improvements in diverse fMRI applications. Deep learning is normally presented as an end-to-end learning process and can alleviate feature engineering requirements and hence reduce domain knowledge requirements to some extent. Under the framework of deep learning, fMRI data can be considered as images, time series or images series. Hence, different deep learning models such as convolutional neural networks, recurrent neural network, or a combination of both, can be developed to process fMRI data for different tasks. In this review, we discussed the basics of deep learning methods and focused on its successful implementations for brain disorder diagnosis based on fMRI images. The goal is to provide a high-level overview of brain disorder diagnosis with fMRI images from the perspective of deep learning applications.

1. Introduction

The functional Magnetic Resonance Imaging (fMRI) is a type of imaging techniques to demonstrate regional, time-varying changes in brain metabolism, and is one of the forms of neuroimaging to infer brain activities by measuring changes in brain blood flow. In fact, it indirectly measures blood-oxygenation-level-dependent (BOLD) signals in human brain, and the BOLD signals are then used to quantify and visualize brain functions. Resting-state fMRI (rs-fMRI) is widely used in neuroimaging tool boxes that measure spontaneous fluctuations across the whole brain without any task being carried on [1–4]. fMRI has had a significant growth since it was born in early 1990s, and become a common tool for clinical and academic uses such as preclinical investigation, neuroscience study, human brain mapping, neural metabolism etc [2]. fMRI is very useful in studying human brain disorders such as Schrizophrenia [5], Autism [6], Alzheimer [7] etc. The study and interpretation of fMRI data is interdisciplinary with roots in neurology, signal processing, machine learning, graph theory and so on. Inspired and enabled by fMRI time series, sophisticated modeling of brain functional networks leads to new understanding of the human brain for different tasks such as brain disorder diagnosis. More recently, attention is paid to the use of machine and statistical learning methods to infer cognitive brain states from fMRI data and build brain networks and so on [8–10], and deep learning as state-of-art learning models is used in fMRI analysis at an accelerated pace.

fMRI comes with relatively lower spatial resolution than structural MRI, but it comes with temporal information. fMRI analysis will also make a significant impact on our understanding of the brain. It may be used to examine the brain's functional anatomy in order to determine the brain connectivity, evaluate the effects of stroke or other disease, or to guide brain treatment [3]. fMRI data has been employed to reflect functional integration of the brain. Alteration in brain functional connectivity (FC) has the potential to provide biomarkers for classifying or predicting brain disorders. fMRI may detect abnormalities within the brain that cannot be found with other imaging techniques, especially when the changes are minor and there is no significant structural changes in general. In combination with deep learning models,

☆ This work is supported by NSERC.
* Corresponding author.
E-mail addresses: wutao.yin@usask.ca (W. Yin), longhai@math.usask.ca (L. Li), fangxiang.wu@usask.ca (F.-X. Wu).

https://doi.org/10.1016/j.neucom.2020.05.113
0925-2312/

Please cite this article as: W. Yin, Li and F.-X. Wu, Deep learning for brain disorder diagnosis based on fMRI images, Neurocomputing, https://doi.org/10.1016/j.neucom.2020.05.113

fMRI analytics are becoming prevalent in many brain studies and even clinical trials.

Machine learning (ML) and Artificial intelligence (AI) have undergone dramatic developments in the past decade and attracted plentiful interest from academia, industries and funding agencies. This major breakthrough is inherently driven by deep artificial neural networks, aka, deep learning [11]. Along with open-source deep learning software frameworks such as Tensorflow [12] and Pytorch [13] and explosion of the available high performance computing infrastructures especially GPUs [14] as well as new network architecture [15–17] and training techniques [18–20], neural networks can be trained much deeper than ever before. These deep models are becoming the state-of-the-art algorithms and outperforming many other methods over a wide spectrum of diverse applications including computer vision, natural language processing and speech recognition etc [21]. At the same time, the proliferation of deep learning reshapes the medical imaging research at a fast pace and this technology is highly sought after. The rapid progress and wide scope of deep learning, accelerated by the large amount of investment from industries and government agencies, has led to a great leap in the entire field of machine learning. Deep learning enabled algorithms are giving exciting results in the medical imaging analysis area, and considered to be a key method for future applications, and have the capacity to deal with medical imaging data sets for accurate and efficient diagnosis. Deep learning is playing an important role in medical imaging analytics such as disease diagnosis, image registration and segmentation [22]. Meanwhile, deep learning based methods can alleviate experts' domain knowledge on feature engineering to some extent by introducing the end-to-end learning framework, and hence increases its popularity in specific medical imaging applications and in turn enables researchers to embrace this technology. Deep learning based algorithms are showing a huge potential and even becoming de facto solutions for many real world medical imaging problems. Deep learning permeates almost the entire field of medical image analysis [23,24].

The motivation of this paper is to provide a review of deep learning for functional MRI analysis. As this research area is very broad and rapidly expanding in recent years, we will not survey the entire deep learning applications, but we provide a comprehensive overview of recent advances and challenges in deep learning applied to topics of brain disorder diagnosis from fMRI images. The rest of this paper is structured as follows. In Section 2, for the sake of completeness, we very briefly introduce the main deep learning techniques that have been used for fMRI analysis throughout the paper. Section 3 describes the application of deep learning in brain disorder diagnosis based on fMRI data. Section 4 discusses open challenges, and concludes the paper with a summary and a high-level outlook for future opportunities.

2. The overview of deep learning methods

In this section, we briefly review deep learning techniques that have been used for fMRI analysis. An excellent tutorial can be found here [18]. With regard to more in-depth coverage on deep learning, please refer to this book [19] for more detailed treatment. Here, we only introduce some essentials that are the most relevant in brain fMRI analysis area.

2.1. Artificial intelligence, machine learning and deep learning

2.1.1. Artificial intelligence

AI refers to a branch of computer science dealing with creation of systems performing or imitating intelligent human behavior. An AI system can be loosely interpreted as an intelligent system that machines can do tasks typically requiring human intelligence. AI is a broad concept that encompasses machine learning, where machines can learn from past experiences (data) and generalize into unseen scenarios without human intervention. Deep learning is a subset of machine learning based on artificial neural networks (ANN), inspired by the biological brain networks and its information processing and exchange [19]. AI systems, as an umbrella terminology, often incorporate machine learning, deep learning, expert system, human prior knowledge, and data analytics and so on that can enable intelligent decision making without human intervention once being setup. Fig. 1 visually demonstrates the relationship of artificial intelligence, machine learning and deep learning as explained above.

2.1.2. Machine learning

Machine learning has a long history with many sub fields and is studied by researchers from different backgrounds [25]. According to how the input data is utilized by the learning algorithms during training, machine learning can be typically categorized as supervised learning, unsupervised Learning, semi-supervised learning and reinforcement learning. Most of machine learning systems used in medical images belong to the class of supervised learning, i.e., the machine is given a set of labeled data, trained to discover the hidden patterns in the labeled data set, and then make predictions on the unseen data sets. However, researchers do use unsupervised learning methods such as clustering [26] and autoencoders [27] for some specific medical imaging tasks. Machine learning models are usually formulated as an optimization problem. An machine learning model then is trained with training data, and fine tuned to make accurate predictions based on the training data. The main objective of training the models is to generalize the learned knowledge so that reasonable predictions can be made on unseen data. The model generalization ability is typically estimated during the training, and benchmarked on separate validation and test data sets to show how well the model perform on unseen data. A particular ML model is usually trained to perform a specific task.

2.1.3. Deep learning

Deep learning is a branch of machine learning which is based on artificial neural networks. It is referred as deep learning in contrast with shallow ANNs. The deep ANNs (DNNs) have more layers than ever before that enable learning hierarchical structures in different granularity and a greater amount of composition of learned functions or learned concepts than conventional machine learning algorithms [19,28]. It is a type of representation learning that discovers a hierarchy of structures in the data. Trained with large amount of data, deep learning shows outstanding performance with enormous model capacity, and performs well on diverse structured and unstructured, and even interconnected data sets. The ever increasing data creation during this digital age is another reason that the capabilities of deep learning grow exponentially in recent years, which reduces the training difficulty in general and helps researchers from different backgrounds access this exciting technology. In addition to more open data sets, deep learning algorithms benefit from stronger computing power available nowadays and allow more participants to get involved.

2.2. Brief introduction to deep learning techniques in fMRI analysis

At the same time, the amount of medical image data generated by health facilities and research institutes are staggering, and could be efficiently utilized by deep learning in the coming years. With the end-to-end training pipeline, deep learning alleviates the requirements of traditional feature engineering process, and allows machine learning researchers to apply know-hows on versatile

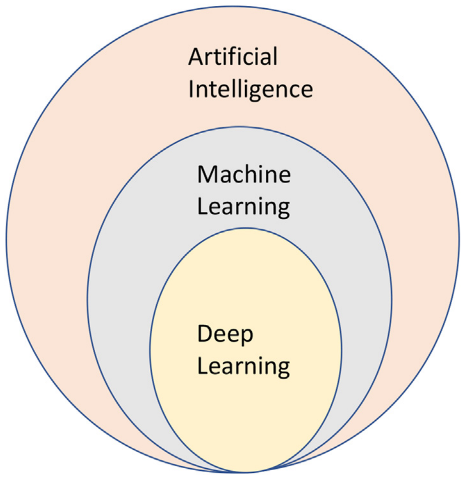


Fig. 1. This figure demonstrates the relationship of artificial intelligence, machine learning and deep learning. AI is an umbrella term with a broad scope. Currently, machine learning is a prevalent technology to empower AI system with intelligence.

subjects. The auto feature discovery capabilities, more or less, reduces domain expert knowledge requirement. Deep learning technology is proliferating to medical imaging research area and receiving bulk of attention from academia and industry, showing great impacts on a diverse subjects from upstream imaging acquisition [29] to downstream disease diagnosis [23,30]. Though the deep learning research in medical images is still at its early stage, there are already a lot of exciting results published [22,31]. Among numerous variants of DNNs, Convolutional neural networks (CNN) and recurrent neural networks (RNN) are most commonly used types of deep learning methods in medical image analysis. Currently, CNNs are most widely used in (medical) image analytics, but RNNs are ramping up quickly, especially when temporal information involved. The following subsections will give an overview of these network architectures with a brief discussion of potential applications and challenges when applied to medical image problems. These neural networks are nowadays typically implemented in one or more prevalent software frameworks such as Tensorflow or Pytorch, and running on top of NVIDIA's GPUs and associated computing libraries. Fig. 2 shows a simplified deep learning based brain disorder diagnosis flowchart that includes four common types of fMRI data representation and utilization methods. According to different perspective, fMRI data can be utilized as 2D/3D images, time series, a combination of both images and time series (aka, 4D images), and functional connectivity network etc. Based on different fMRI data interpretation and representation, different deep learning models can be devised to perform classifications tasks. In general, distinct representation and understanding of fMRI calls for different deep neural networks architectures and requires careful selections of neural network and model training.

2.2.1. Convolutional neural networks

A Convolutional neural network (CNN) is a neural network that has one or more stacking convolutional layers and are used mainly for image processing, classification, segmentation etc [32,33]. The idea is to have a sliding convolution kernel filter over the input with controllable strides, and CNN's model size is controlled by network depth and associated convolutional filter size per layer. Compared to standard feedforward neural networks with similar sizes such as multilayer perceptrons (MLP), CNNs have fewer connections and parameters and are easier to train with relatively less data to achieve satisfactory performance [33]. The attractive feature of CNN is that it can exploit spatial or time correlation of the data efficiently. CNN is also a type of feed-forward hierarchical neural networks and can be divided into multiple learning stages composed of convolutional layers, non-linear activation units, and pooling layers. Stacking linear and non-linear processing units help learn complicated hierarchical representations of the data at different levels of abstraction. Useful features are extracted from locally correlated data points by convolution operators, and fed to nonlinear activation units in order to learn the abstract representation from raw data directly. These two operations enable CNNs to learn feature representations directly from raw sensory data, and reduces feature construction and selection requirements. The most common use for CNNs is image classification, but it is not only limited to this area. They are widely used in medical image processing [34–37]. Fig. 3 shows a simplified CNN network for an image classification task.

2.2.2. Recurrent neural network

Recurrent Neural Networks (RNNs) take the previous output or hidden states as inputs and enable to capture temporal and context features from the data, especially the long-term dependencies such as numerical times series from sensors, stock markets etc [38]. In traditional neural networks, inputs and outputs are considered to be independent of each other, but when it is required to take previous samples into consideration such as machine translations. The most important feature of RNN is its hidden state that can remember some previous information of a sequence, thus RNNs can model sequences of elements that are not independent. The difference between RNNs and other neural networks is that they take both time and sequence into account. That is, RNNs can simultaneously model sequential and time dependencies at multiple scales [38]. Among many variants of RNNs, long short-term memory (LSTM) [39] and gated recurrent units (GRU) [40] are the most well-known subset, and they help solve the vanishing (and/or exploding) gradient problem in the RNN training process and reduce the difficulty of training a recurrent neural network in practice. In the realm of sequence modeling, RNNs have shown its outstanding performance in versatile applications including speech and audio processing, machine translation [41,42,40,43–45]. RNNs are also used in fMRI image analysis such as modelling brain hemodynamic response [46] and exploring temporal dynamics in fMRI time series data [47].

2.2.3. Auto encoder and decoder

An autoencoder (AE) is a type of artificial neural network used to learn hidden representations of the data in an unsupervised manner. The goal of AEs is to minimize the reconstruction error between the input and the output, and learn important features present in the data [27]. AE uses a pair of encoder and decoder structures to create an output that is close to the original input. AEs compress the input into a lower dimension representation and then reconstruct an output from this representation as similar to the input as possible. The latent representation learned by AEs outperforms those handcrafted and predefined features in many applications. It can learn representations of the data and also can be used to reduce the dimensionalitiy of the input data [48]. AEs play an important role in unsupervised learning and transfer learning [49–51]. In fMRI image analysis, AEs are typically used to compress the data and learn the latent representation for further analysis such as classification [52]. Fig. 4 shows a simplified symmetric auto-encoder network.

3. Deep learning in brain disorder diagnosis

One of the most important objectives of neuroimaging research is to find biomarkers that may assist the diagnosis of brain

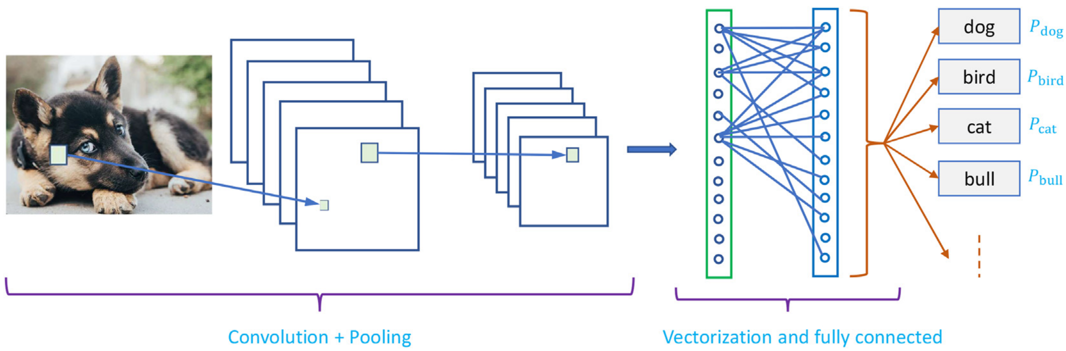


Fig. 2. This figure shows a simplified flowchart for brain disorder diagnosis under a deep learning framework by using different fMRI data interpretations.

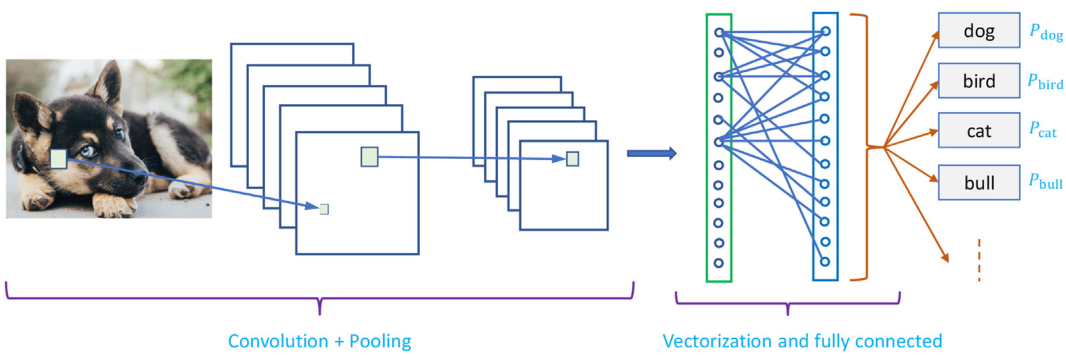


Fig. 3. This figure shows a simplified CNN network architecture for classification purpose.

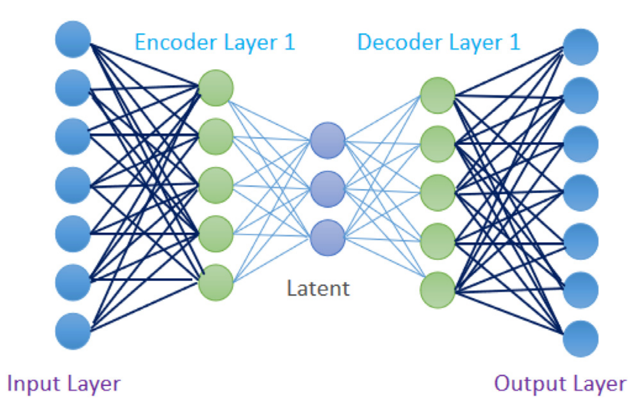


Fig. 4. This figure shows a simplified autoencoder network. Typically, the input layer's size is larger than the latent layer so that the dimension of the input data is reduced.

disorders and/or help treat these disorders. Using deep learning algorithms to investigate neurological malfunctions has the advantage of alleviating feature engineering, and an ever-increasing number of neuroimaging studies are turning to deep learning enabled methods in order to demystify these neurological disorders [53–55,9,56–58]. In this section, we discuss deep learning's applications in brain disorder detection and diagnosis. Deep learning methods are penetrating clinical practice and reshaping medical imaging research community, and research articles involving deep learning are accumulating at a fast pace. Hence, we do not attempt to give an exhaustive review on deep learning, but mainly focus on fMRI images based brain disorder diagnosis and show the landscape of the active research initiatives. There are many different approaches to analyze fMRI images from different perspectives. Features can be extracted from fMRI images to perform disorder classifications. Different feature extraction methods and data analysis tools can result in different deep learning models. In the following, the content is organized according to the types of features being used by deep learning algorithms.

3.1. Functional connectivity model based approaches

3.1.1. Functional connectivity model construction

Before we discuss applications of deep learning to fMRI images using functional connectivity (FC) models and we briefly introduce the FC model first [55,9]. The analysis of functional connectivity in the brain leads to the understanding of the behavior of different brain regions, and resting state functional connectivity (rsFC) as an alternative is also a useful technique for the study of neurological disorders. The human brain connectome has been to delineate the functional sub-regions of the whole brain and understand the functional role of each region of interest (ROI) in a brain cortex. FC approaches to understand the brain from functional networks point of view require an anatomical or data-driven functional atlas to create connections between ROIs and construct a brain functional network for further study. Even though no single functional/anatomical atlas has dominated, an underlying assumption is that such a valid functional atlas exists. This assumption is the foundation for functional connectivity models [2–4].

fMRI images are widely used and accepted to scrutinize the brain activation patterns and their relationships with mental disorders. Alteration in brain functional connectivity is believed to have the potential to locate bio-markers for classifying or predicting neurological disorders. Functional connectivity analyses are used to explore the intrinsic organization of the brain typically at rest or during some other tasks. Similarities in the temporal fluctuations of BOLD signals are measured, and ROIs that share similar temporal patterns are considered as functionally connected. Functional connectivity is an observable phenomenon that can be quantified with mathematical measures such as the ROI based temporal correlation, aka seed-based method, extracted by using data driven methods such as independent component analysis (ICA), dictionary learning etc., or analyzed by using graph theory based methods [59–64]. In common, many functional connectivity networks are static by construction. However, there is extensive neurological evidence showing that functional structure of the human brain varies dynamically as the brain switches among cognitive states even at rest. Functional networks also varies across subjects and tasks. Recently, deviating from the stationary assumption of brain networks, a lot of research and neurological evidences have shown that rsFC demonstrates variations, meaningful dynamics, and repeatable patterns during a typical rs-fMRI scan time [65–68]. Depending on how the correlation coefficients are defined and calculated, the connectivity can be static or dynamic as shown in Fig. 5 (a) and (b). Static connectivity is defined across the whole fMRI scan interval, while the dynamic connectivity is often calculated by using the sliding window with a pre-defined window size.

3.1.2. FC based DL methods

Recently, the efficacy of brain functional connectivity network based classification has been proven by many excellent research works [9,69–71]. Functional connectivity characterizes the brain functional network structure and can be conveniently extracted from fMRI data for classifications [72]. Connectivity based methods are prevalent and straightforward for brain disorder classifications, and a lot of research has been devoted to utilizing functional connectivity as features to train classifiers. Functional connectivity between pairs of ROIs is one of the most common feature representations of fMRI data under the machine learning framework. Fig. 6 shows a simplified machine learning pipeline based on functional connectivity model. The connectivity parameterization step in this machine learning pipeline defines the connectivity matrix extractions. There are many existing work on how to extract connectivity parameterization and in theory they all can be used as input features for deep learning models. Pearson correlation between pairs of ROIs are the most commonly used mathematical measures to define the connectivity strength [72,73], or alternatively functional connectivity matrices constructed by projecting the covariance estimates onto a tangent space [74,72] or recently developed dynamic time warping [75] based method can be used as well.

3.1.3. Direct use of functional connectivity measures

A functional network is composed of ROIs, aka nodes. The connectivity between each node is calculated using the time-series of pair-wise ROIs. One convenient and straightforward approach is to utilize functional connectivity measures out of the connectivity matrix as the feature and directly feed them to a deep learning model, i.e. to perform the fMRI image classification based solely on brain functional activation patterns. This technique is widely applied by a lot of researchers to disease diagnosis due to its simplicity and efficacy. In this part, we review the different deep networks that directly use functional connectivity for disease classification.

Feed-forward neural network, aka, multiple layer perceptron (MLP) is one of the common types of neural network architecture and finds a lot of applications in FC based classification tasks. Campbell et al. [76] reported a deep learning based machine learning method to identify the robust biomarkers for level of pathologically induced consciousness from rs-fMRI data. Combining multiple scale features (i.e., regional homogeneity, local amplitude of low-frequency fluctuations, and functional connectivity based network), a multiple layer perceptron was trained to classify the pathological consciousness level. Santos et al. [77] also reported an MLP based classification model to predict cocaine dependence from fMRI signals as cocaine abusers have shown disrupted/altered functional connectivity while performing a cognitive task or even during resting state. The authors stated that using Pearson correlation coefficients led to the best performance among other correlation measures, and also reported that the MLP consistently performed better than other methods such as support vector machine (SVM), Bayes network etc. Lori et al. [78] summarized the commonly used DL based machine learning pipeline to discover fingerprints from fMRI data in order to identify accurately a specific subject from a large group. On top of over sampling based data augmentation methods, an MLP with two hidden layers was proposed by Eslami et al. [79] for Autism disease (ASD) classifications. In this paper, the training dataset was oversampled to double the size including both healthy controls and ASD subjects for the sake of reducing and/or avoiding overfitting. Yan et al. [80] also applied DNNs for Schizophrenia classification based on functional connectivity network and tried to find out the high-level and relevant feature representation for interpretation purpose. Deshpande et al. [81] proposed a fully connected cascade neural network model for studying attention deficit hyperactivity disorder (ADHD). Instead, the authors derived directional connectivity measures between brain regions from individual subjects and used them as features in classification. It was reported that the performance of DNN classifier was consistently high and stable compared to SVMs for classification of ADHD based don fMRI images.

The convolutional neural network architecture is also one of the frequently used deep models and has found many applications in FC based disease diagnosis. Santana et al. [82] used dynamic time warping distance to define the correlation values between ROI time-series pairs and constructed the connectivity matrix to classify chronic pain conditions. Then one convolutional neural network was trained to distinguish healthy controls from chronic pain patients. This CNN based classifier outperformed traditional machine learning techniques and was not significantly impacted by brain parcelation atlas, and instead the dynamic time warping based connectivity value significantly increased the classifier performance over other correlation measures.

Autoencoders are constantly used to learn latent representations and reduce the dimensions of functional connectivity matrices as well as pursing better generalization of predictive models. In [52], Heinsfeld et al. used deep auto-encoders to reduce the dimensionality of the vectorized on-diagonal and lower-triangle of the connectivity matrix in order to classify subjects as Austism diseases and healthy controls. After knowledge extraction by autoencoders, the encoder weights were fed to an MLP for classification. The auto-encoder architecture showed great potential of coping with the latent factors from intricate structures in the raw data and guide the classification process. The authors achieved the improved performance on a large-scale multi-site dataset, i.e., ABIDE I dataset.[1] The classification across multiple sites is usually difficult due to site differences, scanner configurations and other none controlled variations. The achievement of a reliable classification accuracy showed promise for machine learning applications to clinical datasets, and even possible identification of mental disorders.

Due to its efficacy, a lot of literature used autoencoders to first denoise the functional connectivity matrices before classification. Hu and Ju et al. [83,84] proposed another autoencoder architecture to perform the classification and prediction of Alzheimer disease (AD) patients. The AE was pre-trained and fine tuned before feeding into a softmax classifier for improved accuracy. Mostafa et al. [85] applied autoencoders to extract high level features from raw features based on eigenvalues and centralities of functional brain networks constructed with the entire ABIDEI dataset. Eslami

[1] http://fcon_1000.projects.nitrc.org/indi/abide/abide_I.html

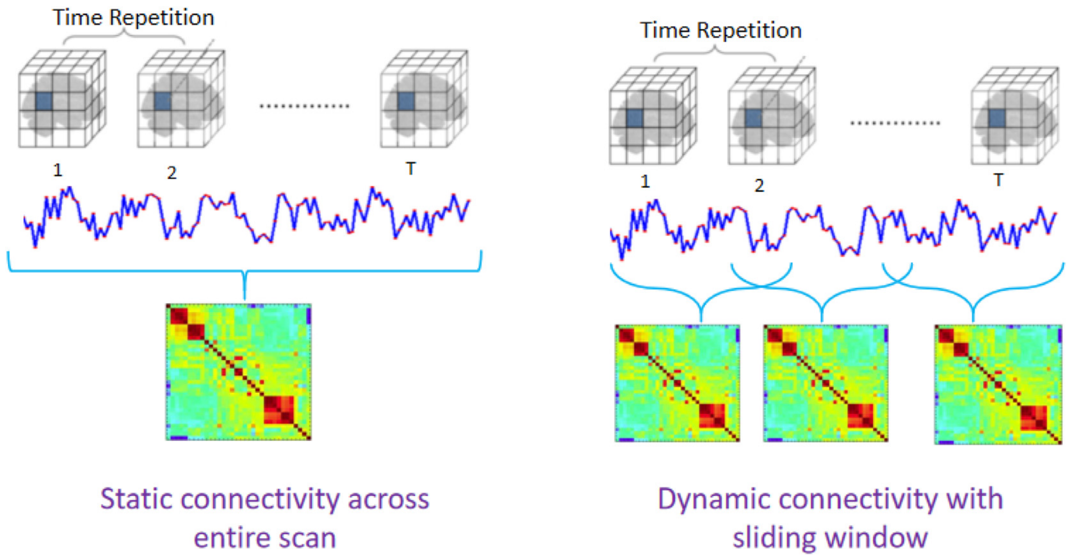


Fig. 5. This figure shows static and dynamic connectivity models derived from fMRI time series data.

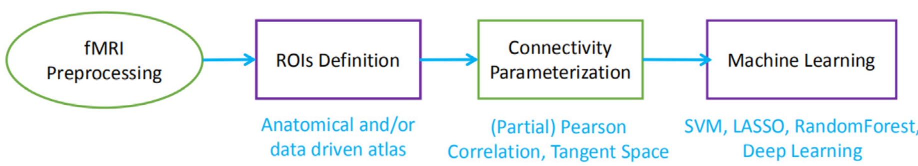


Fig. 6. This figure shows a functional connectivity based machine learning pipeline for brain disorder classification.

et al. [86] also utilized a similar autoencoder architecture in their ASD-DiagNet model, but the authors jointly trained both an autoencoder and a single layer perceptron by combining two loss functions, i.e., mean square error loss for reconstruction, and binary cross entropy for the classification task. Data augmentation techniques based on linear interpolation were also utilized to increase the number of samples for training purpose on the ABIDE I dataset. Kim et al. [5] also proposed a DNN classifier with pre-training and weight sparsity control and demonstrated significantly enhanced performance for automated diagnosis of Schizophrenia patients, and the pre-training weights were from a stacked autoencoder and the pre-trained weight initialization method showed significant performance boost. The authors examined deep hierarchical features extracted by the DNN models and found out that lower (or higher/global) features of the whole-brain functional connectivity patterns were learned in the lower (or higher) layers of the deep neural network. Zeng et al. [87] developed a discriminant autoencoder network with sparsity constraint aimed at learning site-shared rather than site-specific fMRI image features. In contrast to the conventional sparse autoencoder network [5], an optimized discriminant item based on correlation function was introduced in the cost function at the pre-training stage to generate discriminating fMRI features for binary Schizophrenia classification. Guo et al. [88] proposed a DNN model with a novel feature selection method based on multiple autoencoders for predicting ASD from brain resting-state functional connectivity patterns. The feature selection process was used to eliminate redundant features often emerging in the feature layer because many hidden layer nodes had similar activation values. Features with high discriminating power were of interest during the feature selection procedure due to the fact that they can generate more informative representations for the classification model. It is worth noting that Wang et al. [89] devised a SVM-recursive feature elimination algorithm (SVM-RFE) based feature selection method to perform a rough filtering of the primitive functional connectivity features, and then trained a stacked sparse autoencoder with two hidden layers to extract the high-level latent and complicated features from the selected features by SVM-RFE. Finally, the optimal features obtained were fed into the softmax classifier. The authors reported the best classification accuracy for Autism diagnosis on ABIDE I datasets. In [90] Dekhil proposed a sparse autoencoder for ASD diagnosis. However, the features used in diagnosis were the power spectral densities (PSDs) corresponding to the activation time courses of the atlas of interest created by probabilistic ICA. To represent PSDs in a higher level, a sparse autoencoder was used to encode the PSDs through a set of nonlinear filters to a new space. Then SVM classifiers were used to perform the classifications.

3.1.4. End-to-end model for disease classification

It is common to process the time series of ROIs first and construct the functional connectivity, and then apply some classic or deep learning based classification algorithms. However, in [91,92], Riaz et al. proposed an end-to-end deep learning model for the classification of ADHD that takes pre-processed fMRI time-series signals directly as input and predicts healthy control and ADHD patients as output. The proposed work is motivated by FCNet [93] to extract functional connectivity from fMRI time-series signals. FCNet combines CNNs and classical machine learning such as kernel methods and was not trained end-to-end. However, these two papers [91,92] presented a novel end-to-end deep learning model for classification of a neurological disorder from fMRI data. This CNN based network successfully extracted features from individual ROI time-series signals and learned abstract representations of the individual time-series signal. This proved that it is possible to devise an end-to-end deep learning model for the classification of a neurological disorder. The end-to-end approach provided flexibility and may stimulate more new ideas on how to design deep neural network for disease diagnosis.

3.1.5. Connectivity matrices as an analogy to 2D images

In light of success of convolutional neural networks in image classification and object recognition, Meszlenyi et al. [94] proposed a convolutional neural network architecture for functional connectome classification on amnestic mild cognitive impairment disease. The authors argued that the task of classification based on brain functional connectivity data showed remarkable similarities to the image classification, and hence CNN should find its application in functional connectivity classifications. This paper reported remarkable robust results by using CNN based neural networks for fMRI image classifications. Similarly, Sherkatghanad et al. [95] reported an automated detection of autism spectrum disorder using CNN with ABIDE I dataset by applying CNNs to functional connectivity data. The proposed approach was able to classify ASD and control subjects based on the patterns of functional connectivity by concatenation of several convolution layers of different sizes. The combination of CNNs of different sizes was used to learn the representations among different regions of human brains and its associated interaction by using functional connectivity data.

3.1.6. Comment on functional connectivity based methods

The functional connectivity model is one of the most important approaches to handle fMRI data. However, even if the choice of connectivity-based measures as features for brain disorder classification is biologically inspired and effective, noise in the fMRI data has a huge impact on the classification performance. Not all connectivity parameters are relevant for classification and thermal and physiological noise may play a major role. Consequently, feature selections should have a prominent impact on the performance of classifiers. Especially, the sample size of fMRI data is relatively smaller compared to the amount of functional connectivity parameters due to lack of labeled data. Hence, the classification accuracy depends on the discriminative power of the input feature.

The above reviewed references suggest that deep learning methods can classify big multi-site brain image datasets such as ABIDE I. However, in most papers, deep learning based classifiers needs to be in combination with some other feature selection or feature compression methods such as autoencoders. This possibly suggests that the achievement of a reliable classification accuracy requires more well labeled fMRI data to train a relatively deep model as the number of functional connectivity parameters is in general large. Despite this problem, deep learning methods still show bright futures in the assistance of brain disorder diagnosis.

3.2. 2D/3D image processing perspective

fMRI data can be considered as 3D imaging sequences, and as such there are a lot of researchers studying fMRI data from a high dimensional image perspective in light of the remarkable success of CNNs in image classification. It is imperative to study fMRI images using deep learning based advanced image processing techniques. In this subsection, we review the fMRI image classification problem from this point of view and organize this subsection by how the input high-dimensional image-like data is created and used.

3.2.1. 4D fMRI to 2D images conversion

One of the direct image conversion method is to unfold 4D fMRI volumes to a stack of 2D images, and then train the DL network from a 2D image processing point of view. Sarraf et al. [96] proposed a convolutional neural network to distinguish an Alzheimers brain from a healthy one. Using CNNs and the famous architecture LeNet-5 [97], the authors successfully classified Alzheimers subjects from normal controls based on fMRI images. The preprocessed fMRI 4D data was concatenated along Z and time axes and then converted to a stack of 2D images. The proposed deep learning pipeline opened new avenues in fMRI image processing, and explored the inherent spatial features of fMRI data. In [98], Ramzan et al. applied a similar approach to convert preprocessed 4D fMRI scans to 2D images along with image height and time axis. Then the newly created 2D image datasets were used to train a residual network [17] for the early diagnosis of neuro-degenerative disorders such as Alzheimer's disease. Note that the authors also reported in this paper that pre-training and transfer learning approach helped achieve the better performance. Puranik et al. [99] applied ResNet V2 in a similar pipeline as in [98] for AD classifications. In [100], a 2D image conversion method was also employed to convert 4D fMRI volume, and then a CNN based pipeline was designed in order to distinguish different stages of Alzheimer's disease. LeNet-5 [97] and GoogLeNet [101] were both attempted with the excellent performance on a subset of ADNI dataset.[2] Yang et al. [102] applied a similar LeNet-5 as [100] for ASD diagnosis.

In the above work, well-performed networks were typically applied to 2D image stacks converted from 4D fMRI volume without optimization of network hyper-parameters such as number of layers, feature map sizes etc. In [103], an iterative optimization of CNN network method, called Deep Network Optimization, was proposed to select more appropriate network architecture, and this approach was verified on the ADHD classifications. In [104], Farahana et al. proposed an end-to-end CNN learning approach to lay a foundation for basis feature map size selection. The authors studied different feature map sizes and their associated impact on the classification accuracy. It aimed to produce a simple rule of thumb for sizing the feature maps for CNN's application in fMRI classifications.

3.2.2. 3D neural network

fMRI images can be considered as 3D image sequences along the time axis. Hence, many deep learning methods such as CNNs alike can be directly ported to analyze spatial features of fMRI images with modifications to accommodate the 3D features. There were many researchers trying to preserve the inherent 3D spatial locality out of an fMRI volume, and a lot of attempts were conducted to reduce a 4D volume to a 3D image-like data. Then 3D-CNNs alike neural network architectures could be utilized for classifications.

3.2.2.1. 3D CNN and its variations. It was prevalent to consider using a 3D-CNN directly for fMRI image-based classification after accounting for the temporal information. Khosla et al. [9] proposed an alternative representation of connectivity data that leveraged 3D CNNs to build a prediction model that exploited 3D spatial structure of rs-fMRI for ASD classification. An ensemble learning strategy was hence designed to combine predictions from multiple 3D-CNN classifiers corresponding to different brain parcellations. The input to the 3D CNN was constructed by concatenating voxel-level maps of "connectivity fingerprints" and the connectivity maps were represented as a multi-channel 3D volume. This model showed excellent performance by exploring the temporal connectivity information through correlation and also extracting the spatial features at the same time. Similarly, Qureshi et al. [105] also proposed a 3D-CNN based deep learning classification framework, but instead 3D ICA based functional network maps

[2] http://adni.loni.usc.edu/

were used as input of the 3D-CNN model. These 3D images were acquired by performing dual regression on the group ICA results and the constructed ICA functional network maps after dual regression served as discriminative 3D imaging features into a 3D-CNN model for the discrimination of schizophrenia without any thresholding or mask. In [106], Duc et al. also directly applied a 3D-CNN model to spatial maps created by 3D ICA regression for the diagnosis of patients with AD. Similarly in [107], a 3D ICA spatial map approach was used to create 3D volumetric images for AD diagnosis, but with a VGGNet [16] based 3D-CNN network architecture for the feature extraction and classification purpose. Voxel-wise whole brain FC measures were stacked as a multi-channel FC images and these functional connectivity images were used as input to the deep CNNs for building a brain age prediction model [108]. In [109,110], Kam et al. used the same ICA concepts to construct spatial brain functional maps (BFN) and trained multiple 3D CNNs. Each channel of the 3D CNNs learned complex hierarchical spatial pattern information from one BFN in a layer-by-layer manner. Both static and dynamic BFNs can fit into this multi-channel 3D-CNN framework to learn the deeply embedded spatial patterns. These CNNs for multiple BFNs are further combined with consequential layers for end-to-end diagnosis of early mild cognitive impairment. This multiple channel deep learning framework fused high-level features of each single BFN CNN in a unified framework for a joint eMCI classification and showed robust and consistent performance. In [111–113], Li et al. applied a novel sliding-window to move along the time dimension of the 4D fMRI sequence and calculated the mean and standard deviation for each voxel's time series within the sliding window, thus 2-channel (mean and std) 3D image were generated for further processing. Then the authors devised a 2-channel 3D-CNN (2CC3D) to successfully handle high dimensional data and capture the spatial information. Using a sliding window approach, enough data was generated from limited subjects along with fMRI image noise reduction. Experiments showed that the 2CC3D method using mean and std-channels as 3D CNN input performed well for ASD classification. Furthermore, middle layer features of the 2CC3D model showed promise for spatially local information extraction for classification. Vu et al. [114] employed the 3D-CNN for raw 3D fMRI volumes as inputs to do classification. It was argued that the 3D-CNN model with 3D volume input was potentially better suited for fMRI volume classification by automatically extracting shift-invariant features from the fMRI volumes despite of potential issues such as spatial misalignment during normalization and spatial variability of activation patterns across sessions and/or subjects. Inspired by the way radiologists examining brain images, Zou et al. [115] aimed to automatically diagnose ADHD via the 3D CNN using rs-fMRI signals. In their paper, the fMRI volume was first encoded on 3 popular types of 3D low-level features, resembling to three color channels of video data. Then, the 3D CNN was used to explore the topology information to boost the ADHD classification performance.

3.2.2.2. 3D autoencoders. Oh et al. [116] used a General Linear Model (GLM) method to construct 3D activation map based on the contrast images (aka, regressed out the time axis) for Schizophrenia disease diagnosis. The 3D activation maps were first used for pre-training with a 3D-Convolutation AutoEncoder (CAE), and then the encoding layers of the 3D-CAE were selected as 3D-CNN feature extraction layers for supervised fine-tuning. In the proposed deep learning model, 3D-CAE and 3D-CNN were incorporated to devise an end-to-end learning framework without using hand-crafted features. At the same time spatial locality of 3D fMRI was preserved hereby, resulting in more robust feature representation. Their paper reported that the 3D-CAE-based CNN had higher accuracy compared to other models, and visualization of salient brain regions was possible with important local information for clinical usage. Han et al. [117] reported the hierarchical convolutional sparse auto-encoder considering all dimensional information together. The latent feature representation learned by the hierarchical CAE conserved abundant detail information for the neuroimaging classification. The proposed algorithm was verified by testing on three human brain fMRI datasets, and showed great potentials. In [118], the authors also reported that CNNs were used to implement an autoencoder to extract features related to the high activity areas for classification. The extracted feature matrix was fed into various machine learning algorithms for the smoke relapse classification.

3.2.3. Challenges and opportunities

In the above subsection, fMRI is typically treated as 2D/3D image sequences, or 'movie' of MRI images. The 3D spatial locality-representation is sought after as the most prominent discriminative features and deep convolutional neural networks were applied to learn the high-level hidden spatial representations. However, the difference from regular image/video recognition and classification problems is that the fMRI samples are not so accessible. The largest ever fMRI datasets are typically in the magnitude of thousands while a normal CNN could easily have trainable parameters over 1 million. Limited sample size is a major concern in real-world fMRI classifications from a medical image processing perspective. It might be desirable to incorporate more prior knowledge into deep CNNs while striving to accumulate more samples [115]. Most of purely functional connectivity model considered low level correlations as a vector, neglecting the potential 3D local patterns, while 2D/3D CNN based deep networks by using 2D/3D tensors in contrast preserve spatial patterns to some extent and can offer more location-related insights about biomarkers for brain disorder diagnosis. Even with limited training samples, 2D/3D CNN architecture reviewed previously can still detect physiologically meaningful local patterns from fMRI data. It is concluded that the exploration of spatial information out of fMRI data can be fruitful and offer more spatial insights that typical functional connectivity model can not achieve.

3.3. fMRI images as a time series

It is also possible that fMRI can be interpreted as a time-series and hence many deep learning based time-series classification model can be applied in order to learn the temporal patterns. In this subsection, we briefly review the existing work from a time-series classification point of view.

In [91], the end-to-end network could also be considered to directly perform time-series based classification using multi-channel 1D-CNNs. The end-to-end DL framework was completed by a classifier network for ADHD diagnosis right after the multi-channel feature extraction CNNs. Gazar et al. [119] used the original time courses of spontaneous brain activity as features of interest so that its temporal properties were not compromised for ASD classification. In order to retain all the temporal information and yet deal with the high dimension of the data, a portion of "meaningful" voxels was selected and led to time series per ROIs for a given atlas and then mean time-series within each ROI were extracted. The matrix of these time-series were given as input into a 1D CNN with each channel with respect to one ROI. Binary classification was hence performed between controls and patients. In the paper [120], stacked autoencoders based two-stage architecture was proposed for classification of Schizophrenia subjects based on the functional MRI data. The proposed architecture dealt directly with active voxels' time series without converting them into regional mean time series. The time series was first filtered to get rid of inactive or noisy voxels from the data, and then the complete time series of all active voxels were used as input to autoencoders. The proposed methodology provided an excellent accuracy better than the existing methods on the same dataset. Dvornek et al. [121] proposed the use of recurrent neural networks with LSTMs for classification of ASD patients and typical controls directly based on the rs-fMRI mean region time-series. Data augmentation was also utilized to achieve better prediction accuracy. In a subsequent paper, Dvornek et al. [122,123] proposed to use LSTM networks in combination with other information such as phenotype for accurate identification of ASD from rs-fMRI. In [124], Farias et al. used an LSTM network to identify patients with diabetes by extracting the time-series signals of ROIs.

3.4. Joint spatial and temporal feature exploration

So far, we have seen many fMRI-based classification approaches that mostly use functional connectivity or spatial maps as their input without exploring the fully leveraging the inherent spatial–temporal information in fMRI data. Driven by this goal, fMRI images can also be considered as "video" sequences of 3D volumes during the whole scan. This gives rise to the pursuit of both spatial locality and temporal variations for better performance. With this hypothesis, joint spatial–temporal model possibly can provide more insights for brain disorder diagnosis. In this subsection, we discuss this research initiative.

It is natural to consider integrating RNNs and CNNs sequentially in order to explore both spatial and temporal features. Dakka et al. [125] proposed to exploit both spatial and temporal information in the fMRI movie at the whole-brain voxel level for Schizophrenia disease detection. The proposed model used a recurrent convolutional neural network (R-CNN), i.e., a 3-D CNN followed by a sequential neural network with LSTM units. The CNN extracted spatial features out of fMRI data and then these features were fed to the LSTM network. The LSTM architecture helped to learn the temporal dependencies in long time windows. In [126], Bengs et al. proposed a similar 4D spatio-temporal deep learning approach for ASD classification. Compared to [125], GRU units instead of LSTM were used to learn the temporal dependence. In particular, Yan et al. [127] combined the strengths of CNN and RNN models to develop a GRU based Multi-scale RNN model that can directly work on time series fMRI images for classifying brain disorders. This model avoided the correlation analysis of time series data, and multi-scale convolution layers could be complementary to CNN feature extraction as it accounted for different temporal scales of brain activities. Li et al. [7] applied a similar type of sequential concatenation structure of 3D-CNN and LSTM for AD classification. The model demonstrated the capability to utilize the spatial and temporal information simultaneously. Mao et al. [128] considered the structure of rs-fMRI images as time-series 3D frames and applied different 4D CNNs and CNN-LSTM for ADHD detection. This paper introduced data augmentation and data sampling methods to construct relatively large training datasets to avoid overfitting. It was done by sampling one subject's rs-fMRI frames into several relatively short-term pieces with a fixed stride. The public dataset of ADHD-200 Consortium was used to train and validate our method.

It was novel that Wang et al. [47] embedded CNN units into an RNN pipeline to demonstrate that the inclusion of temporal features greatly improved the identification accuracy. The proposed convolutional RNN seamlessly integrated spatial and temporal features for individual identification with rs-fMRI data. The result showed that the conv-RNN network achieved a higher identification accuracy than conventional RNNs, possibly due to spatial features' contribution.

3.5. Other deep learning models and training techniques

There are some miscellaneous deep learning models and training tricks for fMRI image-based classification applications that are not covered in the previous subsections, yet are worth mentioning. We briefly discuss these applications under the context of brain disorder diagnosis.

3.5.1. Graph CNN and its applications

Brain functional networks constructed by fMRI data may show inherent graph structure by nature. A promising approach to harvest both the advantages of graph theoretic and deep learning methods could be graph neural networks (GNNs) [129], which has the potential to learn the spatial representations in the non-euclidean domain. Recently, there are some research works introducing GNNs into fMRI analytics area. We will review some related literature on disease diagnosis in this subsection.

Several researchers [130–133] applied GNN models to the task of classifications. In [133], Arslan applied a vanilla GNN for the classification of sexes based on the brain functional connectivity matrix derived from task fMRI data. In [132], Ktena et al. used the graph representations to model the functional connectivity derived from fMRI data between a set of ROIs, and proposed to learn a graph similarity metric using a siamese graph convolutional neural network. The proposed framework operated in the graph spectral domain to evaluate the similarity between a pair of graphs. Their method demonstrated to perform tasks of classification between matching and non-matching graphs by evaluating the similarity metrics between different brain connectivity networks. Parisot et al. [130,131] leveraged both imaging and non-imaging information such as phenotype information for brain analysis in large populations instead of using time-series correlations directly and achieved the excellent prediction accuracy. Anirudh et al. [134] extended the work in [130] by incorporating ensemble learning techniques to reduce the sensitivity of initial graph structure generations. In [135], a GNN based method was proposed to integrate all the available connectivity, geometric, anatomic information and fMRI related parameters into graphs for deep learning, and the proposed method performed robustly on different atlases and hyper-parameters.

3.5.2. Generative models

It is worth noting that there is some work with the deep generative models to the task of disease classifications in order to alleviate high-dimensional but low sample size dilemma that fMRI data analysis is currently facing. These models are at its early stage, but they show different perspectives on fMRI image-based classifications and hereby we discuss some of the representative work for the completeness of the review.

As argued by Matsubara et al. in [136,137] that a direct classification has risks of overfitting using a small number of high-dimensional samples, and unsupervised feature-extraction such as autoencoders has risks of extracting features of no interest. The authors proposed a deep neural generative model of rs-fMRI data. The deep generative model described the joint distribution of rs-fMRI images, class labels, and remaining frame-wise variability. Then, if the fMRI data of a subject were more likely generated given the class of patients rather than the class of controls, the subject was classified to have the disorder because of Bayes' rule. The authors reported that the generative model showed better classification accuracy on a small-size dataset than the discriminative model for the task of psychiatric disorder diagnosis. In [138], Suk et al. utilized an autoencoders to encode ROI based time series in order to learn the latent representation, and then trained state-space models with hidden Markov models (HMMs) to model the functional dynamics in rs-fMRI images of normal controls and

MCI patients. Then, finally the classification was equivalent to estimate the likelihood of controls and MCI with the trained HMMs.

3.5.3. Transfer learning

One of the major challenge for deep learning models is the lack of labeled data in medical images for clinical practices. Transfer learning is a technique aiming at alleviating the data scarcity problem in the target domain (e.g., medical imaging analytics) by utilizing the knowledge acquired in the source domain (e.g., image classification). We hereby discuss some representative work with transfer learning techniques.

Inspired by the transfer learning strategy, the authors [139,140] applied a transfer learning method enhancing the classification of whole brain functional connectivity patterns. The authors trained a stacked sparse autoencoder model to perform high-level feature extraction of functional connectivity patterns for the early prediction of cognitive deficits and the diagnosis of ASD. The hypothesis was that healthy functional connectivity patterns learned from an existing large scale database could be transferred to enhance a new brain disorder classification task. The functional connectivity patterns acquired by training on a database of healthy subjects could be transferred to facilitate the application of deep learning models for smaller fMRI datasets. This could be potentially useful for some rare disease detection because the data collection in this scenario would be even more challenging. The authors also reported that the classification performance was significantly improved irrespective of site sample size. In [98], transfer learning was also utilized to train a ResNet [17] from a 2D image processing perspective for eMCI classification. In [99], Puranick also applied transfer learning techniques to train a CNN model for Alzheimer disease diagnosis.

3.6. Summary

In this section, we have discussed different applications of deep learning models and its corresponding applications in fMRI based brain disorder diagnosis. From different perspectives such as brain functional connectivity, image processing, time series, joint spatial–temporal models, various deep learning models and network architecture found their applications in fMRI analytics. Even though deep learning models in fMRI analytics are still at its infancy, they have showed promising results and more often than non deep learning based algorithms reported the state-of-art performance compared to other machine learning methods. With more high-quality fMRI data and innovative neural network architectures becoming available, deep learning models may be seen higher adoption rate and better performance improvement in brain disorder automatic screening/diagnosis, and even play major roles in clinical settings.

4. Challenges and future outlook

4.1. Discussions

Different deep learning models such CNN, RNN, GNN, AE, and transfer learning lead to different classification performance even on the same fMRI dataset. Since fMRI data is very complicated with many features, in general there is no single deep learning model that can fit all tasks. CNN can mostly be used to extract spatial features of fMRI images efficiently, and RNN is better positioned to explore the temporal characteristics of fMRI time series. Sometimes CNNs are also used to perform time series classifications. In this unique setting, fMRI data shows both spatial and temporal features and this motivates to combine both CNNs and RNNs in order to fully explore the potentials of fMRI data. GNNs are more useful to explore the non-Euclidean metrics that exist in a graph-like network, and it is more suitable to learn the hidden representations of functional connectivity network derived from fMRI time series from geometric learning point of view. AE and transfer learning are more frequently used to deal with data scarcity issues. However, AEs are more used to compress data dimensionality by learning the hidden representations of the input data, while transfer learning takes the pre-trained network from similar tasks with abundant samples and migrates the well trained network parameters to new tasks (in this case, it is fMRI data processing). Both AE and transfer learning are used to reduce the fMRI samples and labelling which are typically expensive but from different perspectives. These deep learning models all have their own advantages and disadvantages, and it is imperative to take the merits of different network architectures and make informed decisions when choosing deep learning models for brain disorder diagnosis tasks.

Analysis of clinical medical image data of brains has always been challenging, and the most problematic aspect is selecting the discriminative features with some explainable biomarkers. Neural image analysis such as the fMRI classification has become a heated topic because of its potential to diagnose various brain disorders. Deep learning emerges as state-of-art machine learning techniques and has been favored by many researchers and practitioners. It has become imperative to reap the benefits of deep learning for the fMRI image classifications. There are some obvious advantages of deep learning models.

- A deep learning based fMRI classification pipeline typically has the practical advantage of not requiring the manual feature selection for neurological disorders. This has been an advantage to attract machine learning researchers to study this subject and stimulate more interactions with neurologists and MRI technical experts.
- Compared to shallow models such as kernel machines, deep neural networks have the potential to learn hierarchical representations of the high dimensional fMRI data and discover the inherently complex interactions of the brain functional networks.
- Many successful cases in other domains such as image classifications can be imitated in medical image analysis area due to the similarities, and hence many mature techniques can be transferred to medical image analysis.

However, at the current stage, the fMRI image classification via deep learning models faces many challenges. Successful applications of deep learning to fMRI requires a complete understanding of how these models work and how an appropriate network architecture can be designed for a given task. There are some crucial aspects that should be addressed:

- Deep Learning models have enormous capacity with a huge number of parameters that need to be trained during the learning process. fMRI data is in general of high dimension with a small sample size, and the use of deep learning in small data sets still remains a big challenge. The fMRI data acquisition is typically time consuming and costly. Research projects are under strict scrutiny and regulations and sometimes it slows down the knowledge sharing and transfer process. The accurate classification may benefit from homogeneous datasets of similar scanners and scanning parameters. Many publicly accessible datasets such as ABIDE datasets were generated from different sites with different settings and hence a unified data acquisition method may be highly desirable.
- From the clinical perspective, identifying biomarkers would be more crucial to differentiate brain disorders. Models with explainable powers are easily accepted in a clinical environment. Deep learning models transform the raw fMRI data to

the space of higher level abstraction. This poses challenges to back-map the learned representation to original neural-image space for better interpretations. Some deep learning enabled classification methods may be lack of underlying neural anatomical or functional evidence to back them up.

Although these issues remains not satisfactorily resolved, deep learning methods show a great potential to improve the diagnosis of neural disorders and are considered to be excellent tools for advancing the diagnostic technology. It is gradually becoming a de facto technology for many challenging machine learning tasks in medical imaging analysis.

4.2. Future outlook

Even though great success has been seen in the brain disorder diagnosis with fMRI images in recent years, it is still far away from clinical diagnosis requirement. The prediction accuracy as well as reliable and explainable biomarkers still remain the key focus of the future research. Generalization and interpretability should be kept in mind while developing any predictive models for clinical diagnosis. Most existing methods can only classify one single disorder against healthy control, and intelligent system that can perform multiple disorder classification would be highly sought-after. Incorporating fMRI images with other information such as electronic medical record, EEG, structural MRI image etc [141,142] can possibly yield even better results, and this proposition is still staying as another promising topic.

CRediT authorship contribution statement

Wutao Yin: Conceptualization, Investigation, Visualization, Writing - original draft. **Longhai Li:** Writing - review & editing. **Fang-Xiang Wu:** Conceptualization, Investigation, Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work is supported in part by Natural Science and Engineering Research Council of Canada (NSERC).

References

[1] M.A. Ebrahimighahnavieh, S. Luo, R. Chiong, Deep learning to detect alzheimer's disease from neuroimaging: A systematic literature review, Computer Methods and Programs in Biomedicine 187 (2020) 105242.
[2] N.K. Logothetis, What we can do and what we cannot do with fmri, Nature 453 (2008) 869–878.
[3] G.H. Glover, Overview of functional magnetic resonance imaging, Neurosurgery Clinics of North America 22 (2011) 133–139. Functional Imaging.
[4] L. Loued-Khenissi, O. Döll, K. Preuschoff, An overview of functional magnetic resonance imaging techniques for organizational research, Organizational Research Methods 22 (2019) 17–45.
[5] J. Kim, V.D. Calhoun, E. Shim, J.-H. 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 124 (2016) 127–146.
[6] O. Dekhil, H. Hajjdiab, A. Shalaby, M.T. Ali, B. Ayinde, A. Switala, A. Elshamekh, M. Ghazal, R. Keynton, G. Barnes, A. El-Baz, Using resting state functional mri to build a personalized autism diagnosis system, PloS one 13 (2018), e0206351–e0206351.
[7] W. Li, X. Lin, X. Chen, Detecting alzheimer's disease based on 4d fmri: An exploration under deep learning framework, Neurocomputing (2020).
[8] M.P. van den Heuvel, H.E.H. Pol, Exploring the brain network: A review on resting-state fmri functional connectivity, European Neuropsychopharmacology 20 (2010) 519–534.
[9] M. Khosla, K. Jamison, G.H. Ngo, A. Kuceyeski, M.R. Sabuncu, Machine learning in resting-state fmri analysis, Magnetic Resonance Imaging 64 (2019) 101–121. Artificial Intelligence in MRI.
[10] S.M. Smith, The future of fmri connectivity, NeuroImage 62 (2012) 1257–1266. 20 YEARS OF fMRI.
[11] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436–444.
[12] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G.S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng, TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL: https://www.tensorflow.org/, software available from tensorflow.org.
[13] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, Pytorch: An imperative style, high-performance deep learning library, in: H. Wallach, H. Larochelle, A. Beygelzimer, F. d' Alché-Buc, E. Fox, R. Garnett (Eds.), Advances in Neural Information Processing Systems 32, Curran Associates Inc, 2019, pp. 8024–8035. URL: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.
[14] S. Chetlur, C. Woolley, P. Vandermersch, J. Cohen, J. Tran, B. Catanzaro, E. Shelhamer, cudnn: Efficient primitives for deep learning., CoRR abs/1410.0759 (2014).
[15] N. Zeng, Z. Wang, B. Zineddin, Y. Li, M. Du, L. Xiao, X. Liu, T. Young, Image-based quantitative analysis of gold immunochromatographic strip via cellular neural network approach, IEEE Transactions on Medical Imaging 33 (2014) 1129–1136.
[16] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014. URL: http://arxiv.org/abs/1409.1556, cite arxiv:1409.1556.
[17] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2016 (2016) 770–778, https://doi.org/10.1109/CVPR.2016.90.
[18] G. Hinton, Y. Bengio, Y. LeCun, Deep Learning (2015), https://media.nips.cc/Conferences/2015/tutorialslides/DL-Tutorial-NIPS2015.pdf, nIPS'2015 Tutorial.
[19] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016, http://www.deeplearningbook.org.
[20] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift., CoRR abs/1502.03167 (2015).
[21] M.Z. Alom, T.M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M.S. Nasrin, M. Hasan, B.C. Van Essen, A.A.S. Awwal, V.K. Asari, A state-of-the-art survey on deep learning theory and architectures, Electronics 8 (2019).
[22] A.S. Lundervold, A. Lundervold, An overview of deep learning in medical imaging focusing on mri, Zeitschrift für Medizinische Physik 29 (2019) 102–127. Special Issue: Deep Learning in Medical Physics.
[23] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A. van der Laak, B. van Ginneken, C.I. Sánchez, A survey on deep learning in medical image analysis, Medical Image Analysis 42 (2017) 60–88.
[24] G. Hinton, Deep learning–a technology with the potential to transform health care, JAMA 320 (2018) 1101–1102.
[25] K.P. Murphy, Machine Learning: A Probabilistic Perspective, MIT Press, Cambridge, Mass. [u.a.], 2013. URL: https://www.amazon.com/Machine-Learning-Probabilistic-Perspective-Computation/dp/0262018020/ref=sr_1_2?ie=UTF8&qid=1336857747&sr=8-2.
[26] L. Tang, S. Mostafa, B. Liao, F.-X. Wu, A network clustering based feature selection strategy for classifying autism spectrum disorder, BMC Med Genomics 12 (2019).
[27] G. Dong, G. Liao, H. Liu, G. Kuang, A review of the autoencoder and its variants: A comparative perspective from target recognition in synthetic-aperture radar images, IEEE Geoscience and Remote Sensing Magazine 6 (2018) 44–68.
[28] J. Schmidhuber, Deep learning in neural networks: An overview, Neural Networks 61 (2015) 85–117.
[29] G. Zhu, B. Jiang, L. Tong, Y. Xie, G. Zaharchuk, M. Wintermark, Applications of deep learning to neuro-imaging techniques, Frontiers in Neurology 10 (2019) 869.
[30] N. Zeng, Z. Wang, H. Zhang, K. Kim, Y. Li, X. Liu, An improved particle filter with a novel hybrid proposal distribution for quantitative analysis of gold immunochromatographic strips, IEEE Transactions on Nanotechnology 18 (2019) 819–829.
[31] D. Shen, G. Wu, H.-I. Suk, Deep learning in medical image analysis, Annual Review of Biomedical Engineering 19 (2017) 221–248, PMID:28301734.
[32] Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, L.D. Jackel, Backpropagation applied to handwritten zip code recognition, Neural Computation 1 (1989) 541–551.
[33] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: F. Pereira, C.J.C. Burges, L. Bottou, K.Q. Weinberger (Eds.), Advances in Neural Information Processing Systems 25,

Curran Associates Inc, 2012, pp. 1097–1105. URL: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.

[34] R. Yamashita, M. Nishio, R. Kinh Gian Do, K. Togashi, Convolutional neural networks: an overview and application in radiology, Insights into Imaging 9 (2018) 611–629.

[35] S. Soffer, A. Ben-Cohen, O. Shimon, M.M. Amitai, H. Greenspan, E. Klang, Convolutional neural networks for radiologic images: A radiologist's guide, Radiology 290 (2019) 590–606.

[36] A. Maier, C. Syben, T. Lasser, C. Riess, A gentle introduction to deep learning in medical image processing, Zeitschrift für Medizinische Physik 29 (2019) 86–101. Special Issue: Deep Learning in Medical Physics.

[37] D. Wen, Z. Wei, Y. Zhou, G. Li, X. Zhang, W. Han, Deep learning methods to process fmri data and their application in the diagnosis of cognitive impairment: A brief overview and our opinion, Frontiers in Neuroinformatics 12 (2018) 23.

[38] Z. Lipton, A critical review of recurrent neural networks for sequence learning, arxiv (2015).

[39] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Computation 9 (1997) 1735–1780.

[40] K. Cho, B. van Merriënboer, Ç. Gülçehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using rnn encoder–decoder for statistical machine translation, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Doha, Qatar, 2014, pp. 1724–1734, URL: http://www.aclweb.org/anthology/D14-1179.

[41] J. Donahue, L.A. Hendricks, M. Rohrbach, S. Venugopalan, S. Guadarrama, K. Saenko, T. Darrell, Long-term recurrent convolutional networks for visual recognition and description, IEEE Transactions on Pattern Analysis and Machine Intelligence 39 (2017) 677–691.

[42] S. Venugopalan, H. Xu, J. Donahue, M. Rohrbach, R. Mooney, K. Saenko, Translating videos to natural language using deep recurrent neural networks, in: Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, Denver, Colorado, 2015, pp. 1494–1504, URL: https://www.aclweb.org/anthology/N15-1173. 10.3115/v1/N15-1173.

[43] G. Trigeorgis, F. Ringeval, R. Brueckner, E. Marchi, M.A. Nicolaou, B. Schuller, S. Zafeiriou, Adieu features? end-to-end speech emotion recognition using a deep convolutional recurrent network, in: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2016, pp. 5200–5204, https://doi.org/10.1109/ICASSP.2016.7472669.

[44] H. Zen, H. Sak, Unidirectional long short-term memory recurrent neural network with recurrent output layer for low-latency speech synthesis, in: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2015, pp. 4470–4474, https://doi.org/10.1109/ICASSP.2015.7178816.

[45] H. Salehinejad, S. Sankar, J. Barfett, E. Colak, S. Valaee, Recent advances in recurrent neural networks, 2017. URL: http://arxiv.org/abs/1801.01078, cite arxiv:1801.01078. Comment: arXiv admin note: text overlap with arXiv:1602.04335.

[46] U. Güçlü, M.A.J. van Gerven, Modeling the dynamics of human brain activity with recurrent neural networks, Frontiers in Computational Neuroscience 11 (2017) 7.

[47] L. Wang, K. Li, X. Chen, X.P. Hu, Application of convolutional recurrent neural network for individual recognition based on resting state fmri data, Frontiers in Neuroscience 13 (2019) 434.

[48] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (2006) 504–507.

[49] P. Baldi, Autoencoders, unsupervised learning, and deep architectures., in: I. Guyon, G. Dror, V. Lemaire, G.W. Taylor, D.L. Silver (Eds.), ICML Unsupervised and Transfer Learning, volume 27 of JMLR Proceedings, JMLR.org, 2012, pp. 37–50. URL: http://dblp.uni-trier.de/db/journals/jmlr/jmlrp27.html#Baldi12.

[50] K. Han, H. Wen, J. Shi, K.-H. Lu, Y. Zhang, D. Fu, Z. Liu, Variational autoencoder: An unsupervised model for encoding and decoding fmri activity in visual cortex, NeuroImage 198 (2019) 125–136.

[51] M. Svanera, M. Savardi, S. Benini, A. Signoroni, G. Raz, T. Hendler, L. Muckli, R. Goebel, G. Valente, Transfer learning of deep neural network representations for fmri decoding, Journal of Neuroscience Methods 328 (2019) 108319.

[52] A.S. Heinsfeld, A.R. Franco, R.C. Craddock, A. Buchweitz, F. Meneguzzi, Identification of autism spectrum disorder using deep learning and the abide dataset, NeuroImage: Clinical 17 (2018) 16–23.

[53] S.M. Plis, D.R. Hjelm, R. Salakhutdinov, E.A. Allen, H.J. Bockholt, J.D. Long, H.J. Johnson, J.S. Paulsen, J.A. Turner, V.D. Calhoun, Deep learning for neuroimaging: a validation study, Frontiers in Neuroscience 8 (2014) 229.

[54] S. Vieira, W.H. Pinaya, A. Mechelli, Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications, Neuroscience & Biobehavioral Reviews 74 (2017) 58–75.

[55] G. Martí-Juan, G. Sanroma-Guell, G. Piella, A survey on machine and statistical learning for longitudinal analysis of neuroimaging data in alzheimer's disease, Computer Methods and Programs in Biomedicine 189 (2020) 105348.

[56] T. Jo, K. Nho, A.J. Saykin, Deep learning in alzheimer's disease: Diagnostic classification and prognostic prediction using neuroimaging data, Frontiers in Aging Neuroscience 11 (2019) 220.

[57] G. Zhu, B. Jiang, L. Tong, Y. Xie, G. Zaharchuk, M. Wintermark, Applications of deep learning to neuro-imaging techniques, Frontiers in Neurology 10 (2019) 869.

[58] N. Zeng, H. Qiu, Z. Wang, W. Liu, H. Zhang, Y. Li, A new switching-delayed-pso-based optimized svm algorithm for diagnosis of alzheimer's disease, Neurocomputing 320 (2018) 195–202.

[59] S.E. Joel, B.S. Caffo, P.C.M. van Zijl, J.J. Pekar, On the relationship between seed-based and ica-based measures of functional connectivity, Magnetic Resonance in Medicine 66 (2011) 644–657. 21394769[pmid].

[60] K.A. Smitha, K. Akhil Raja, K.M. Arun, P.G. Rajesh, B. Thomas, T.R. Kapilamoorthy, C. Kesavadas, Resting state fmri: A review on methods in resting state connectivity analysis and resting state networks, The Neuroradiology Journal 30 (2017) 305–317. 28353416[pmid].

[61] E. Bullmore, O. Sporns, Complex brain networks: graph theoretical analysis of structural and functional systems, Nature Reviews Neuroscience 10 (2009) 186–198.

[62] Y.-B. Lee, J. Lee, S. Tak, K. Lee, D.L. Na, S.W. Seo, Y. Jeong, J.C. Ye, Sparse spm: Group sparse-dictionary learning in spm framework for resting-state functional connectivity mri analysis, NeuroImage 125 (2016) 1032–1045.

[63] S. Jeong, X. Li, J. Yang, Q. Li, V. Tarokh, Dictionary learning and sparse coding-based denoising for high-resolution task functional connectivity mri analysis (2017) 45–52, https://doi.org/10.1007/978-3-319-67389-9_6.

[64] G. Varoquaux, A. Gramfort, F. Pedregosa, V. Michel, B. Thirion, Multi-subject dictionary learning to segment an atlas of brain spontaneous activity, in: G. Székely, H.K. Hahn (Eds.), Information Processing in Medical Imaging, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 562–573.

[65] E. Damaraju, E. Allen, A. Belger, J. Ford, S. McEwen, D. Mathalon, B. Mueller, G. Pearlson, S. Potkin, A. Preda, J. Turner, J. Vaidya, T. van Erp, V. Calhoun, Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia, NeuroImage: Clinical 5 (2014) 298–308.

[66] E.A. Allen, E. Damaraju, S.M. Plis, E.B. Erhardt, T. Eichele, V.D. Calhoun, Tracking whole-brain connectivity dynamics in the resting state, Cerebral Cortex (New York, N.Y.) 24 (2014) (1991) 663–676.

[67] R.M. Hutchison, T. Womelsdorf, E.A. Allen, P.A. Bandettini, V.D. Calhoun, M. Corbetta, S. Della Penna, J.H. Duyn, G.H. Glover, J. Gonzalez-Castillo, D.A. Handwerker, S. Keilholz, V. Kiviniemi, D.A. Leopold, F. de Pasquale, O. Sporns, M. Walter, C. Chang, Dynamic functional connectivity: promise, issues, and interpretations, NeuroImage 80 (2013) 360–378.

[68] H. Guo, S. Mostafa, W. Yin, F.-X. Wu, Diagnosis of asd from rs-fmris based on brain dynamic networks, in: submitted to 2020 ISBRA, 2019.

[69] Y. Du, Z. Fu, V.D. Calhoun, Classification and prediction of brain disorders using functional connectivity: Promising but challenging, Frontiers in Neuroscience 12 (2018) 525.

[70] A. Abraham, M.P. Milham, A.D. Martino, R.C. Craddock, D. Samaras, B. Thirion, G. Varoquaux, Deriving reproducible biomarkers from multi-site resting-state data: An autism-based example, NeuroImage 147 (2017) 736–745.

[71] T. He, R. Kong, A.J. Holmes, M. Nguyen, M.R. Sabuncu, S.B. Eickhoff, D. Bzdok, J. Feng, B.T. Yeo, Deep neural networks and kernel regression achieve comparable accuracies for functional connectivity prediction of behavior and demographics, NeuroImage 206 (2020) 116276.

[72] K. Dadi, M. Rahim, A. Abraham, D. Chyzhyk, M. Milham, B. Thirion, G. Varoquaux, Benchmarking functional connectome-based predictive models for resting-state fmri, NeuroImage 192 (2019) 115–134.

[73] S. Mostafa, L. Tang, F.-X. Wu, Diagnosis of autism spectrum disorder based on eigenvalues of brain networks, IEEE Access 7 (2019) 128474–128486.

[74] B. Ng, M. Dressler, G. Varoquaux, J.B. Poline, M. Greicius, B. Thirion, Transport on riemannian manifold for functional connectivity-based classification, in: P. Golland, N. Hata, C. Barillot, J. Hornegger, R. Howe (Eds.), Medical image computing and computer-assisted intervention – MICCAI 2014, Springer International Publishing, Cham, 2014, pp. 405–412.

[75] R.J. Meszlényi, P. Hermann, K. Buza, V. Gál, Z. Vidnyánszky, Resting state fmri functional connectivity analysis using dynamic time warping, Frontiers in Neuroscience 11 (2017) 75.

[76] J.M. Campbell, Z. Huang, J. Zhang, X. Wu, P. Qin, G. Northoff, G.A. Mashour, A. G. Hudetz, Pharmacologically informed machine learning approach for identifying pathological states of unconsciousness via resting-state fmri, NeuroImage 206 (2020) 116316.

[77] J.S. Santos, R.M. Savii, J.S. Ide, C.-S.R. Li, M.G. Quiles, M.P. Basgalupp, Classification of cocaine dependents from fmri data using cluster-based stratification and deep learning, in: O. Gervasi, B. Murgante, S. Misra, G. Borruso, C.M. Torre, A.M.A. Rocha, D. Taniar, B.O. Apduhan, E. Stankova, A. Cuzzocrea (Eds.), Computational Science and Its Applications – ICCSA 2017, Springer International Publishing, Cham, 2017, pp. 298–313.

[78] N.F. Lori, I. Ramalhosa, P. Marques, V. Alves, Deep learning based pipeline for fingerprinting using brain functional mri connectivity data, Procedia Computer Science 141 (2018) 539–544. The 9th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN-2018)/ The 8th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH-2018)/ Affiliated Workshops.

[79] T. Eslami, F. Saeed, Auto-asd-network: A technique based on deep learning and support vector machines for diagnosing autism spectrum disorder using fmri data, in: Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, BCB '19, Association for Computing Machinery, New York, NY, USA, 2019, pp. 646–651, https://doi.org/10.1145/3307339.3343482.

[80] W. Yan, S. Plis, V.D. Calhoun, S. Liu, R. Jiang, T. Jiang, J. Sui, Discriminating schizophrenia from normal controls using resting state functional network connectivity: A deep neural network and layer-wise relevance propagation method, in: 2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP), 2017, pp. 1–6, https://doi.org/10.1109/MLSP.2017.8168179.

[81] G. Deshpande, P. Wang, D. Rangaprakash, B. Wilamowski, Fully connected cascade artificial neural network architecture for attention deficit hyperactivity disorder classification from functional magnetic resonance imaging data, IEEE Transactions on Cybernetics 45 (2015) 2668–2679.

[82] A.N. Santana, I. Cifre, C.N. de Santana, P. Montoya, Using deep learning and resting-state fmri to classify chronic pain conditions, Frontiers in Neuroscience 13 (2019) 1313.

[83] C. Hu, R. Ju, Y. Shen, P. Zhou, Q. Li, Clinical decision support for alzheimer's disease based on deep learning and brain network, IEEE International Conference on Communications (ICC) 2016 (2016) 1–6, https://doi.org/10.1109/ICC.2016.7510831.

[84] R. Ju, C. Hu, P. Zhou, Q. Li, Early diagnosis of alzheimer's disease based on resting-state brain networks and deep learning, IEEE/ACM Transactions on Computational Biology and Bioinformatics 16 (2019) 244–257.

[85] S. Mostafa, W. Yin, F.-X. Wu, Autoencoder based methods for diagnosis of autism spectrum disorder, in: to appear in 2019 IEEE ICCABS, 2019.

[86] T. Eslami, V. Mirjalili, A. Fong, A.R. Laird, F. Saeed, Asd-diagnet: A hybrid learning approach for detection of autism spectrum disorder using fmri data, Frontiers in Neuroinformatics 13 (2019) 70.

[87] L.-L. Zeng, H. Wang, P. Hu, B. Yang, W. Pu, H. Shen, X. Chen, Z. Liu, H. Yin, Q. Tan, K. Wang, D. Hu, Multi-site diagnostic classification of schizophrenia using discriminant deep learning with functional connectivity mri, EBioMedicine 30 (2018) 74–85.

[88] X. Guo, K.C. Dominick, A.A. Minai, H. Li, C.A. Erickson, L.J. Lu, Diagnosing autism spectrum disorder from brain resting-state functional connectivity patterns using a deep neural network with a novel feature selection method, Frontiers in Neuroscience 11 (2017) 460.

[89] C. Wang, Z. Xiao, B. Wang, J. Wu, Identification of autism based on svm-rfe and stacked sparse auto-encoder, IEEE Access 7 (2019) 118030–118036.

[90] O. Dekhil, H. Hajjdiab, A. Shalaby, M.T. Ali, B. Ayinde, A. Switala, A. Elshamekh, M. Ghazal, R. Keynton, G. Barnes, A. El-Baz, Using resting state functional mri to build a personalized autism diagnosis system, PloS one 13 (2018), e0206351–e0206351.

[91] A. Riaz, M. Asad, S.M.M.R.A. Arif, E. Alonso, D. Dima, P. Corr, G. Slabaugh, Deep fmri: An end-to-end deep network for classification of fmri data, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), 2018, pp. 1419–1422, https://doi.org/10.1109/ISBI.2018.8363838.

[92] A. Riaz, M. Asad, E. Alonso, G. Slabaugh, Deepfmri: End-to-end deep learning for functional connectivity and classification of adhd using fmri, Journal of Neuroscience Methods 335 (2020) 108506.

[93] A. Riaz, M. Asad, S.M.M.R. Al-Arif, E. Alonso, D. Dima, P. Corr, G. Slabaugh, Fcnet: A convolutional neural network for calculating functional connectivity from functional mri, in: G. Wu, P. Laurienti, L. Bonilha, B.C. Munsell (Eds.), Connectomics in NeuroImaging, Springer International Publishing, Cham, 2017, pp. 70–78.

[94] R.J. Meszlényi, K. Buza, Z. Vidnyánszky, Resting state fmri functional connectivity-based classification using a convolutional neural network architecture, Frontiers in Neuroinformatics 11 (2017) 61.

[95] Z. Sherkatghanad, M. Akhondzadeh, S. Salari, M. Zomorodi-Moghadam, M. Abdar, U.R. Acharya, R. Khosrowabadi, V. Salari, Automated detection of autism spectrum disorder using a convolutional neural network, Frontiers in Neuroscience 13 (2020) 1325.

[96] S. Sarraf, G. Tofighi, Deep learning-based pipeline to recognize alzheimer's disease using fmri data, Future Technologies Conference (FTC) 2016 (2016) 816–820, https://doi.org/10.1109/FTC.2016.7821697.

[97] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE 86 (1998) 2278–2324.

[98] F. Ramzan, M.U.G. Khan, A. Rehmat, S. Iqbal, T. Saba, A. Rehman, Z. Mehmood, A deep learning approach for automated diagnosis and multi-class classification of alzheimer's disease stages using resting-state fmri and residual neural networks, Journal of Medical Systems 44 (2019) 37.

[99] M. Puranik, H. Shah, K. Shah, S. Bagul, Intelligent alzheimer's detector using deep learning, Second International Conference on Intelligent Computing and Control Systems (ICICCS) 2018 (2018) 318–323, https://doi.org/10.1109/ICCONS.2018.8663065.

[100] Y. Kazemi, S. Houghten, A deep learning pipeline to classify different stages of alzheimer's disease from fmri data, IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB) 2018 (2018) 1–8, https://doi.org/10.1109/CIBCB.2018.8404980.

[101] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Computer Vision and Pattern Recognition (CVPR), 2015, http://arxiv.org/abs/1409.4842.

[102] S. Yang, Xin Sarraf, N. Zhang, Deep learning-based framework for autism functional mri image classification, Journal of the Arkansas Academy of Science 72 (2018).

[103] A.M.S. Aradhya, A. Ashfahani, Deep network optimization for rs-fmri classification, International Conference on Data Mining Workshops (ICDMW) 2019 (2019) 77–82, https://doi.org/10.1109/ICDMW.2019.00022.

[104] M.S. Nur Farahana, Z. Htike, Feature map size selection for fmri classification on end-to-end deep convolutional neural networks, International Journal of Advanced and Applied Sciences 5 (2018).

[105] M.N.I. Qureshi, J. Oh, B. Lee, 3d-cnn based discrimination of schizophrenia using resting-state fmri, Artificial Intelligence in Medicine 98 (2019) 10–17.

[106] N.T. Duc, S. Ryu, M.N.I. Qureshi, M. Choi, K.H. Lee, B. Lee, 3d-deep learning based automatic diagnosis of alzheimer's disease with joint mmse prediction using resting-state fmri, Neuroinformatics 18 (2020) 71–86.

[107] M.N.I. Qureshi, S. Ryu, J. Song, K.H. Lee, B. Lee, Evaluation of functional decline in alzheimer's dementia using 3d deep learning and group ica for rs-fmri measurements, Frontiers in Aging Neuroscience 11 (2019) 8.

[108] H. Li, T.D. Satterthwaite, Y. Fan, Brain age prediction based on resting-state functional connectivity patterns using convolutional neural networks, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), 2018, pp. 101–104, https://doi.org/10.1109/ISBI.2018.8363532.

[109] T.-E. Kam, H. Zhang, D. Shen, A novel deep learning framework on brain functional networks for early mci diagnosis, in: A.F. Frangi, J.A. Schnabel, C. Davatzikos, C. Alberola-López, G. Fichtinger (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2018, Springer International Publishing, Cham, 2018, pp. 293–301.

[110] T. Kam, H. Zhang, Z. Jiao, D. Shen, Deep learning of static and dynamic brain functional networks for early mci detection, IEEE Transactions on Medical Imaging 39 (2020) 478–487.

[111] X. Li, Asd biomarker detection on fmri images: Feature learning with data corruptions by analyzing deep neural network classifier outcomes, Yale University Poster, 2018.

[112] X. Li, N.C. Dvornek, X. Papademetris, J. Zhuang, L.H. Staib, P. Ventola, J.S. Duncan, 2-channel convolutional 3d deep neural network (2cc3d) for fmri analysis: Asd classification and feature learning, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), 2018, pp. 1252–1255, https://doi.org/10.1109/ISBI.2018.8363798.

[113] X. Li, N.C. Dvornek, J. Zhuang, P. Ventola, J.S. Duncan, Brain biomarker interpretation in asd using deep learning and fmri, in: A.F. Frangi, J.A. Schnabel, C. Davatzikos, C. Alberola-López, G. Fichtinger (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2018, Springer International Publishing, Cham, 2018, pp. 206–214.

[114] H. Vu, H. Kim, J. Lee, 3d convolutional neural network for feature extraction and classification of fmri volumes, 2018 International Workshop on Pattern Recognition in Neuroimaging (PRNI) (2018) 1–4, https://doi.org/10.1109/PRNI.2018.8423964.

[115] L. Zou, J. Zheng, M.J. McKeown, Deep learning based automatic diagnoses of attention deficit hyperactive disorder, IEEE Global Conference on Signal and Information Processing (GlobalSIP) 2017 (2017) 962–966, https://doi.org/10.1109/GlobalSIP.2017.8309103.

[116] K. Oh, W. Kim, G. Shen, Y. Piao, N.-I. Kang, I.-S. Oh, Y.C. Chung, Classification of schizophrenia and normal controls using 3d convolutional neural network and outcome visualization, Schizophrenia Research 212 (2019) 186–195.

[117] X. Han, Y. Zhong, L. He, P.S. Yu, L. Zhang, The unsupervised hierarchical convolutional sparse auto-encoder for neuroimaging data classification, in: Y. Guo, K.J. Friston, A.A. Faisal, S.L. Hill, H. Peng (Eds.), BIH, volume 9250 of Lecture Notes in Computer Science, Springer, 2015, pp. 156–166, URL: http://dblp.uni-trier.de/db/conf/brain/brain2015.html#HanZHYZ15.

[118] A. Tahmassebi, A. Gandomi, I. McCann, M. Schulte, A. Goudriaan, A. Meyer-Base, Deep learning in medical imaging: fmri big data analysis via convolutional neural networks (2018) 1–4, https://doi.org/10.1145/3219104.3229250.

[119] A. Gazzar, L. Cerliani, G. van Wingen, R. Thomas, Simple 1-d convolutional networks for resting-state fmri based classification in autism, 2019, pp. 1–6. doi:10.1109/IJCNN.2019.8852002.

[120] P. Patel, P. Aggarwal, A. Gupta, Classification of schizophrenia versus normal subjects using deep learning, in: Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing, ICVGIP '16, Association for Computing Machinery, New York, NY, USA, 2016. URL: https://doi.org/10.1145/3009977.3010050. doi:10.1145/3009977.3010050.

[121] N.C. Dvornek, P. Ventola, K.A. Pelphrey, J.S. Duncan, Identifying autism from resting-state fmri using long short-term memory networks, Machine Learning in Medical Imaging. MLMI (Workshop) 10541 (2017) 362–370. 29104967[pmid].

[122] N.C. Dvornek, P. Ventola, J.S. Duncan, Combining phenotypic and resting-state fmri data for autism classification with recurrent neural networks, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), 2018, pp. 725–728, https://doi.org/10.1109/ISBI.2018.8363676.

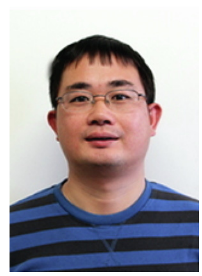
[123] N.C. Dvornek, D. Yang, P. Ventola, J.S. Duncan, Learning generalizable recurrent neural networks from small task-fmri datasets, in: A.F. Frangi, J.A. Schnabel, C. Davatzikos, C. Alberola-López, G. Fichtinger (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2018, Springer International Publishing, Cham, 2018, pp. 329–337.

[124] A.F.S. Farías, A. Mendizabal, A.A. González-Garrido, R. Romo-Vázquez, A. Morales, Long short-term memory neural networks for identifying type 1 diabetes patients with functional magnetic resonance imaging, 2018 IEEE Latin American Conference on Computational Intelligence (LA-CCI) (2018) 1–4, https://doi.org/10.1109/LA-CCI.2018.8625231.

[125] J. Dakka, P. Bashivan, M. Gheiratmand, I. Rish, S. Jha, R. Greiner, Learning neural markers of schizophrenia disorder using recurrent neural networks (2017).

[126] M. Bengs, N. Gessert, A. Schlaefer, 4d spatio-temporal deep learning with 4d fmri data for autism spectrum disorder classification, 2019.
[127] W. Yan, V. Calhoun, M. Song, Y. Cui, H. Yan, S. Liu, L. Fan, N. Zuo, Z. Yang, K. Xu, J. Yan, L. Lv, J. Chen, Y. Chen, H. Guo, P. Li, L. Lu, P. Wan, H. Wang, H. Wang, Y. Yang, H. Zhang, D. Zhang, T. Jiang, J. Sui, Discriminating schizophrenia using recurrent neural network applied on time courses of multi-site fmri data, EBioMedicine 47 (2019) 543–552.
[128] Z. Mao, Y. Su, G. Xu, X. Wang, Y. Huang, W. Yue, L. Sun, N. Xiong, Spatio-temporal deep learning method for adhd fmri classification, Information Sciences 499 (2019) 1–11.
[129] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, P. Yu, A comprehensive survey on graph neural networks, 2019. URL: https://arxiv.org/pdf/1901.00596.pdf.
[130] S. Parisot, S.I. Ktena, E. Ferrante, M. Lee, R. Guerrero, B. Glocker, D. Rueckert, Disease prediction using graph convolutional networks: Application to autism spectrum disorder and alzheimer's disease, Medical Image Analysis 48 (2018) 117–130.
[131] S. Parisot, S.I. Ktena, E. Ferrante, M. Lee, R.G. Moreno, B. Glocker, D. Rueckert, Spectral graph convolutions for population-based disease prediction, in: M. Descoteaux, L. Maier-Hein, A. Franz, P. Jannin, D.L. Collins, S. Duchesne (Eds.), Medical Image Computing and Computer Assisted Intervention MICCAI 2017, Springer International Publishing, Cham, 2017, pp. 177–185.
[132] S.I. Ktena, S. Parisot, E. Ferrante, M. Rajchl, M. Lee, B. Glocker, D. Rueckert, Metric learning with spectral graph convolutions on brain connectivity networks, NeuroImage 169 (2018) 431–442.
[133] S. Arslan, S.I. Ktena, B. Glocker, D. Rueckert, Graph saliency maps through spectral convolutional networks: Application to sex classification with brain connectivity, in: D. Stoyanov, Z. Taylor, E. Ferrante, A.V. Dalca, A. Martel, L. Maier-Hein, S. Parisot, A. Sotiras, B. Papiez, M.R. Sabuncu, L. Shen (Eds.), Graphs in Biomedical Image Analysis and Integrating Medical Imaging and Non-Imaging Modalities, Springer International Publishing, Cham, 2018, pp. 3–13.
[134] R. Anirudh, J.J. Thiagarajan, Bootstrapping graph convolutional neural networks for autism spectrum disorder classification (2017).
[135] X. Li, N.C. Dvornek, Y. Zhou, J. Zhuang, P. Ventola, J.S. Duncan, Graph neural network for interpreting task-fmri biomarkers, in: D. Shen, T. Liu, T.M. Peters, L.H. Staib, C. Essert, S. Zhou, P.-T. Yap, A. Khan (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, Springer International Publishing, Cham, 2019, pp. 485–493.
[136] T. Matsubara, T. Tashiro, K. Uehara, Deep neural generative model of functional mri images for psychiatric disorder diagnosis, IEEE Transactions on Biomedical Engineering 66 (2019) 2768–2779.
[137] K. Kusano, T. Tashiro, T. Matsubara, K. Uehara, Deep generative state-space modeling of fmri images for psychiatric disorder diagnosis (2019) 1–7, https://doi.org/10.1109/IJCNN.2019.8852448.
[138] H.-I. Suk, C.-Y. Wee, S.-W. Lee, D. Shen, State-space model with deep learning for functional dynamics estimation in resting-state fmri, NeuroImage 129 (2016) 292–307.
[139] H. Li, N.A. Parikh, L. He, A novel transfer learning approach to enhance deep neural network classification of brain functional connectomes, Frontiers in Neuroscience 12 (2018) 491.
[140] L. He, H. Li, S.K. Holland, W. Yuan, M. Altaye, N.A. Parikh, Early prediction of cognitive deficits in very preterm infants using functional connectome data in an artificial neural network framework, NeuroImage: Clinical 18 (2018) 290–297.
[141] S.M. Plis, M.F. Amin, A. Chekroud, D. Hjelm, E. Damaraju, H.J. Lee, J.R. Bustillo, K. Cho, G.D. Pearlson, V.D. Calhoun, Reading the (functional) writing on the (structural) wall: Multimodal fusion of brain structure and function via a deep neural network based translation approach reveals novel impairments in schizophrenia, NeuroImage 181 (2018) 734–747.
[142] Y. Hao, H.M. Khoo, N. von Ellenrieder, N. Zazubovits, J. Gotman, Deepied: An epileptic discharge detector for eeg-fmri based on deep learning, NeuroImage: Clinical 17 (2018) 962–975.

Wutao Yin received the BSc degree and MSc degree from Northwestern Polytechnical University and University of Saskatchewan in 2006 and 2009 respectively. He is now pursuing a PhD degree in the Division of Biomedical Engineering at the University of Saskatchewan. His research interest lies in medical imaging analysis, biomedical signal processing, bioinformatics, machine learning, deep learning and artificial intelligence.



Longhai Li received his Ph.D. degree in statistics from the University of Toronto in 2007. He received his B.Sc in statistics from the University of Science and Technology of China in 2002. He is currently a full professor at the University of Saskatchewan. His research activities focus on developing and applying statistical machine learning methods for high-throughput data and complex-structured data.

Fang-Xiang Wu received both the BSc and MSc degrees in applied mathematics from Dalian University of Technology in 1993 and 1990, respectively, the Ph.D. in bioinformatics and computational biology from the University of Saskatchewan (U of S) in August 2004 and another Ph.D. in control theory and its applications from Northwestern Polytechnical University in 1998 He is currently a full professor in College of Engineering and Department of Computer Science at the U of S. His research interests include on artificial intelligence, machine/deep learning, computational biology and bioinformatics, image analytics, complex network analytics and controllability.