Magnetic Resonance Imaging 64 (2019) 171–189

Contents lists available at ScienceDirect

Magnetic Resonance Imaging

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

Review article

Role of deep learning in infant brain MRI analysis

Mahmoud Mostapha[a], Martin Styner[a,b,*]

[a] Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, United States of America
[b] Neuro Image Research and Analysis Lab, Department of Psychiatry, University of North Carolina at Chapel Hill, NC 27599, United States of America

ARTICLE INFO

Keywords:
MRI
Infant MRI
Machine learning
Deep learning
Convolutional neural networks
Isointense segmentation
Prediction

ABSTRACT

Deep learning algorithms and in particular convolutional networks have shown tremendous success in medical image analysis applications, though relatively few methods have been applied to infant MRI data due numerous inherent challenges such as inhomogenous tissue appearance across the image, considerable image intensity variability across the first year of life, and a low signal to noise setting. This paper presents methods addressing these challenges in two selected applications, specifically infant brain tissue segmentation at the isointense stage and presymptomatic disease prediction in neurodevelopmental disorders. Corresponding methods are reviewed and compared, and open issues are identified, namely low data size restrictions, class imbalance problems, and lack of interpretation of the resulting deep learning solutions. We discuss how existing solutions can be adapted to approach these issues as well as how generative models seem to be a particularly strong contender to address them.

1. Introduction

Neurodevelopmental disorders (NDDs) are a diverse group of conditions that are characterized by deficiencies in brain functions including cognition, communication, behavior and motor skills resulting from atypical brain development. Attention-deficit/hyperactivity disorder (ADHD), autism spectrum disorder (ASD) and intellectual disability are examples of disorders that fall under the umbrella of NDD [1]. Such conditions can be chronically disabling resulting in significant long-term impairment to the adult life of those children who suffered them. Due to the lack of biomarkers to diagnose NDD or to differentiate among them, these disorders are still currently identified based on clinical interviews and behavioral observations. Therefore, NDDs share common needs that include (*1*) the development of a clinically-useful, early presymptomatic test for identifying infants who will develop neurodevelopmental disorders, which will enable more efficient early intervention in infancy, (*2*) the bijective assessment of disease severity through inspecting associated dysfunctional brain systems and contrasting them to typically functional ones, thereby leading to accurate long-term disease prognosis that will help to accelerate the evaluation of potentially applied interventions.

Advances in Magnetic Resonance Imaging (MRI) enabled the non-invasive visualization of the infant's brain through acquired high-resolution images. The increasing availability of large-scale datasets of detailed infant brain multi-modal MR images, e.g., T1-weighted (T1w), T2-weighted (T2w), diffusion-weighted MRI (dMRI), and resting-state functional MR (rsfMRI) images, affords unique opportunities to accurately study early postnatal brain development leading to insights into the origins and abnormal developmental trajectories of NDDs. However, the processing and analysis of MR infant brain images are typically far more challenging as compared to the adult brain setting. As illustrated in Fig. 1, an infant's brain MRI suffers from reduced tissue contrast, large within-tissue inhomogeneities, regionally-heterogeneous image appearance, large age related intensity changes and severe partial volume effect due to the small brain size. Since most of the existing tools were designed for adult brain MRI data, infant-specific computational neuroanatomy tools are a relatively recent development. As shown in Fig. 2, a typical pipeline for early prediction of NDDs from infant structural MRI (sMRI) consist of three main steps. (*1*) Image preprocessing, tissue segmentation, and regional labeling, and extraction of image-based features [2–4]. (*2*) Surface reconstruction, surface correspondence, surface parcellation, and extraction of surface-based features [5–7]. (*3*) Feature preprocessing, feature extraction, machine learning model training, and prediction of unseen subjects [8–11].

As acquiring such multi-modal MR infant brain images becomes more common, the collections of medical imaging data available to researchers are increasing in number, size, and complexity. For instance, the Baby Connectome Project (BCP) is aiming to acquire

* Corresponding author at: Neuro Image Research and Analysis Lab, Department of Psychiatry, University of North Carolina at Chapel Hill, NC 27599, United States of America.
E-mail addresses: mahmoudm@cs.unc.edu (M. Mostapha), styner@unc.edu (M. Styner).

https://doi.org/10.1016/j.mri.2019.06.009
Received 15 January 2019; Received in revised form 6 June 2019; Accepted 8 June 2019

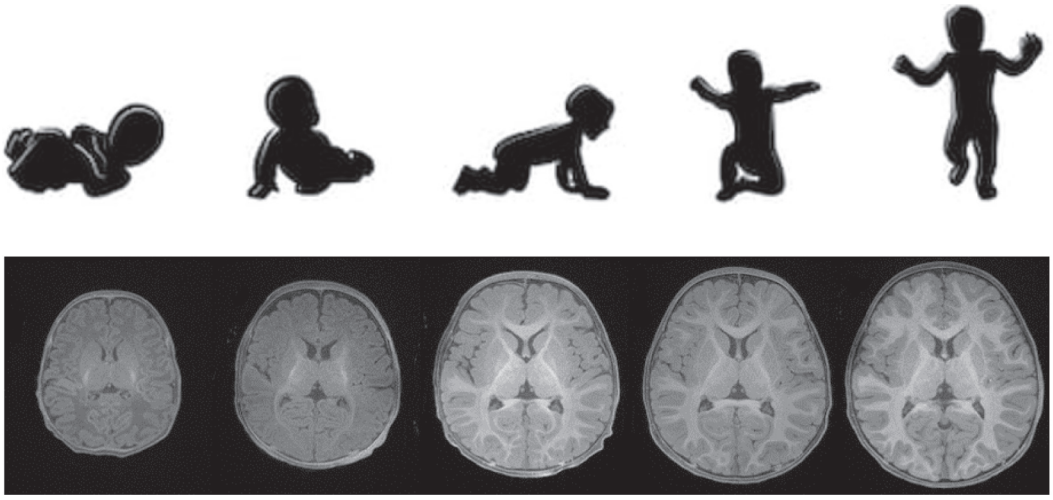


Fig. 1. T1w images of a typically-developing infant, scanned longitudinally at 0, 3, 6, 12 and 24 months of age.

longitudinal high-resolution multi-modal MRI data from > 500 typically-developing children from birth to 5 years of age [12]. Hence, this necessitates the development of methods that can utilize information extracted from these large datasets. In the realm of big data, artificial intelligence (AI) and machine learning are often used interchangeably. However, AI has a broader notion than machine learning, which tackles the use of computers to mimic the humans cognitive functions. As illustrated in Fig. 3, machine learning is a subset of AI that focuses on the design and development of algorithmic techniques that allow computer systems to evolve function or behavior by learning from large data. Classical machine learning algorithms have a variety of applications that have been tailored to the medical imaging field [13–15]. However, most necessitate the creation of "hand crafted" image features through careful engineering and specific domain expertise. Also, such traditional machine learning algorithms tend to not generalize well to new, previously unseen data [16]. This problem is amplified in medical imaging applications due to the inherent anatomical variability in brain morphology, discrepancies in acquisition settings and MRI scanners, and variations in the appearance of pathological tissues. In recent years, a machine learning technique referred to as deep learning [16,17] has gained wide popularity in many artificial intelligence applications due to its ability to overcome limitations of classical machine learning algorithms.

Deep learning is sometimes referred to as "deep neural networks," referring to the many layers involved compared to traditional neural networks that may only have a single layer of data (Fig. 4). Such a deep network architecture enables the extraction of a complex hierarchy of features from input data via self-learning as opposed to the engineered feature extraction in classical machine learning algorithms. Deep learning shows impressive performance and generalizability through training on a large amount of data [18–20]. This success is due mostly to the rapid progress in computational power, in particular through graphics processing units (GPUs), which enabled the fast development of complex deep learning algorithms. Several types of deep learning architectures have been developed for different tasks including object detection, speech recognition, and classification mainly in computer vision. In turn, the success of deep learning in computer vision lead to its use in medical image analysis, for image segmentation [21], image registration [22], image fusion [23], lesion detection [24], and computer-aided diagnosis [25]. In this paper, we provide an example based overview of recently developed deep learning techniques with applications to the field of infant brain MR segmentation and NDDs prediction. Also, we will discuss some of the remaining performance gaps that have the potential to be fulfilled by new advancements in the field of deep learning.

2. Deep learning for infant brain tissue segmentation

Accurate segmentation of infant MRI into regions of different tissue classes, such as white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF), is a crucial early processing step in extracting imaging biomarkers [26,27]. Due to the ongoing myelination and maturation processes in the first year of life [28], we can roughly distinguish four stages of distinct appearance of the brain in MR images [29], namely, (1) a neonate/infantile stage where WM-GM contrast is inverse as compared to adults (≤ 4 months), (*2*) an isointense stage (5–9 months) when contrast between WM and GM is low or absent, and (*3*) an early adult-like stage (≥ 10 months) when WM-GM contrast appears similar to the adult stage (*4*). As shown in Fig. 1, at the neonate stage (0–3 months) and at adult-like stages, T1w and T2w MRI typically show reasonable contrast that enables accurate separation of different structure, though at lower signal to noise ratio (SNR) and contrast-to-noise ratio (CNR) at the neonate stage.

2.1. Brain tissue segmentation at neonatal stage

The segmentation of a neonatal brain (≤ 1 month) is challenging due to the lower SNR and CNR caused by the shorter scanning time imposed by expected motion constraints and the small size of the neonatal brain. Moreover, the CSF-GM boundary has a similar intensity profile to the mostly unmyelinated WM leading to severe partial volume (PV) effects. Additionally, the large variability caused by the rapid brain development and the ongoing WM myelination process impose additional challenges in developing accurate segmentation tools [30]. Over the years, several non-deep learning based methods were proposed to accurately segment neonatal brains, which can be mainly categorized into parametric [31–33], classification [34,35], multi-atlas fusion [36,37], deformable models [38,39]. Neonatal tissue segmentation methods were evaluated in the NeoBrainS12 2012 MICCAI Grand Challenge (http://neobrains12.isi.uu.nl), where T1w and T2w images were provided with corresponding manually segmented images. Most methods showed accurate segmentation results, where classification-based approaches showed highly accurate results but more sensitive compared to other methods. However, the segmentation of myelinated vs unmyelinated WM remains a challenge as most methods failed to produce accurate results consistently [30]. This can be attributed

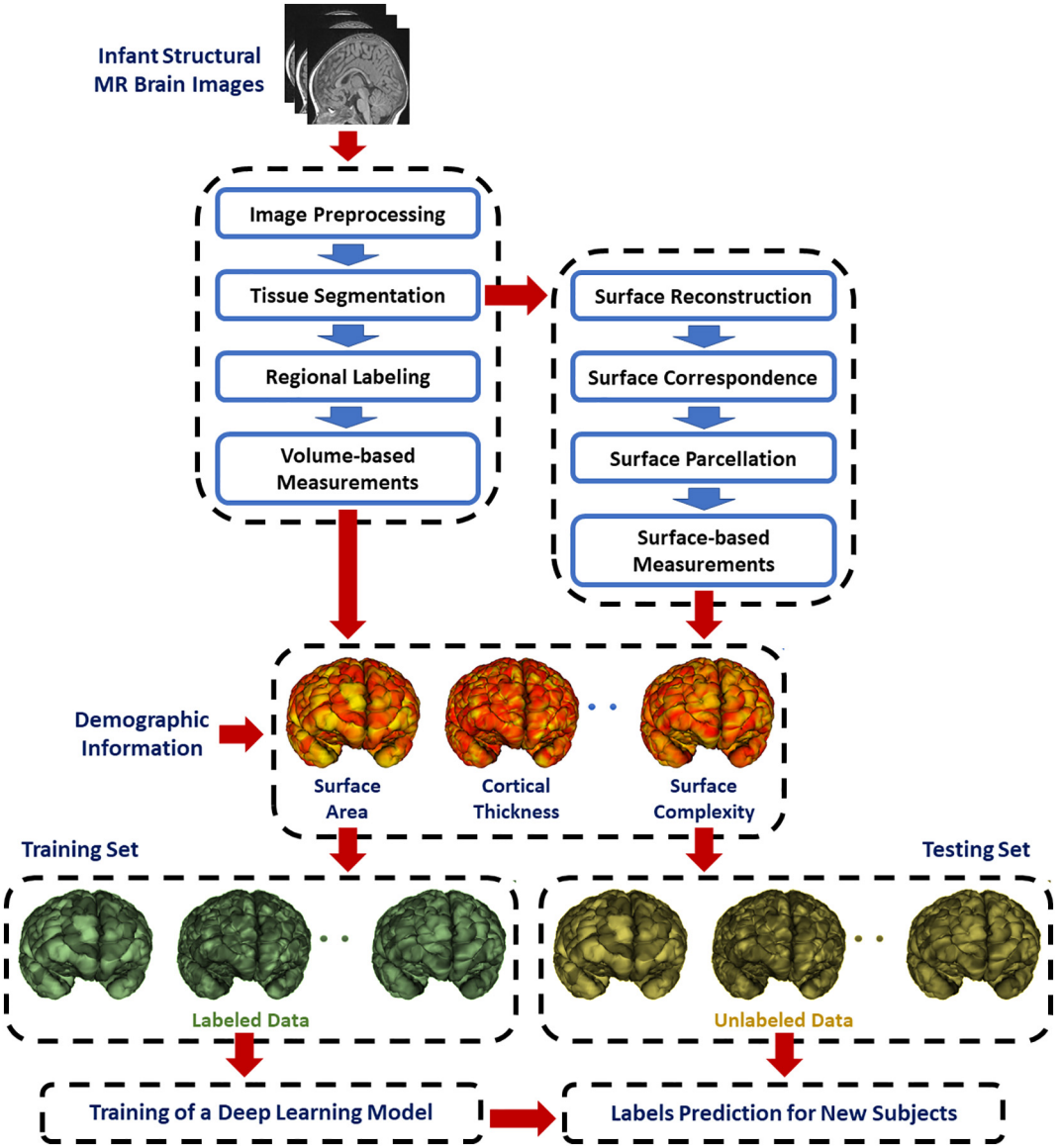


Fig. 2. A typical pipeline for a machine learning approach for early prediction of neurodevelopmental disorders (NDDs) using infant structural MR brain images (sMRI). Please note that this workflow employs infant-specific processing and analysis steps.

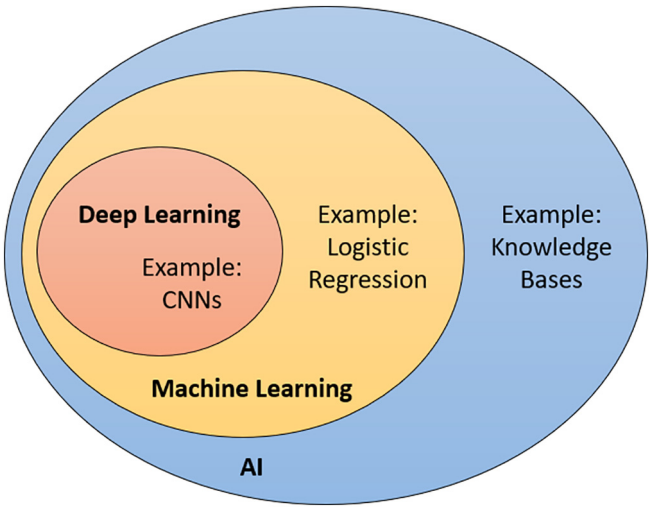


Fig. 3. Venn diagram illustrating the relationship of deep learning, machine learning and artificial intelligence.

mainly to the lack of accurate manually segmented atlases that describe these regions efficiently. To improve the current segmentation results, deep learning techniques particularly using convolutional neural networks (CNNs) are being investigated. However, only a few studies (Moeskops et al. [40] and Rajchl et al. [41]) attempted to utilize CNNs for such task, which can also be attributed to the lack of adequate, accurate manually segmented data needed to train deep learning models. Ongoing efforts through projects such as BCP and dHCP are promising to solve this problem by enabling the availability of publicly available annotated datasets.

2.2. Brain tissue segmentation at isointense stage

Another time point where traditional segmentation methods are still challenged is around six months of age. At this age stage, both T1w and T2w scans show a high overlap between the intensity distributions of both WM and GM (isointense), which imposes a significant challenge for accurate tissue segmentation [38]. In addition to the extremely low tissue contrast between WM and GM, isointense infant MRI images still suffer from a generally lower signal-to-noise and higher partial volume

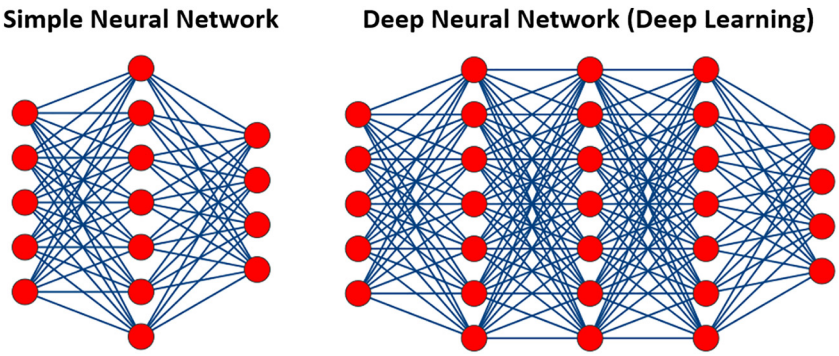


Fig. 4. Architectures of two feed-forward fully-connected neural networks. A classical neural network containing one hidden layer and a deep neural network (deep learning) that has two or more hidden layers. Modern deep neural networks can be based on tens to hundreds of hidden layers.

effects as compared to adult scans. Due to the necessity to acquire isointense infant MRI scans during nature sleep of the subject, additional complicating factors affecting image quality such as short acquisition times and increased motion artifacts are present. Finally, biological factors such as spatially variable myelination rates across the brain further complicate the segmentation of isointense infant brain MRI data. In this paper, we will focus on this particularly difficult task of accurately segmenting infant MRI brain images at the isointense stage as methods developed for this time point have high potential for successful translation to other time points including the neonatal stage. To solve this problem, many traditional segmentation methods have been proposed. Atlas-based approaches [38,42–45] incorporate a-priori knowledge about the location of different brain structures through the use of a single atlas or more commonly multiple atlases. In the multi-atlas setting, atlases are first deformably registered to the subject image, and labels are then combined using a label fusion strategy. In addition to tissue segmentation, such methods are commonly also applied for the segmentation of subcortical structures [46]. Atlas-based segmentation techniques show reasonable accuracy in many applications. Nevertheless, they are still challenged by registration errors caused by low contrast and intensity inhomogeneities present in the infant population. A variety of techniques have been proposed as a refinement step to overcome some of the limitations associated with atlas-based approaches. Probabilistic methods [47,48] will formulate the segmentation as an energy minimization problem in which shape or spatial priors constrain voxel probabilities. On the other hand, deformable-based models [49] will refine contours produced by atlas-based approaches to better match computed image edges. These refinement methods typically require large labeled datasets that generally are not available and are harder to generalize to multi-tissue segmentation.

Longitudinal approaches have also been proposed where images of the same subjects scanned at an older, adult-like stage are employed via longitudinal co-registration [11,50]. Such methods rely on the availability of follow up scans which might not be practical. Recently, alternative have been proposed that do not need such follow up data, such as the anatomy-guided joint tissue segmentation and topological correction framework for isointense infant MRI proposed by Li [51]. This approach adaptively integrates information from both intensity images and estimated tissue probability maps, as well as employs prior anatomical information in the form of a signed distance map to the outer cortical surface to correct topological errors. Although the results obtained by this method show reasonable accuracy compared to previous work, the method is still limited by choice of anatomical priors and the reliance on engineered, potentially non-optimal discriminative features that are harder to generalize to more massive datasets. Therefore, there is a need for techniques that can automatically learn general features from the infant data in a data-driven manner. Recently, manually labeled six months data has become available through different resources including segmentation challenges such as the iseg 2017 MICCAI Grand Challenge (http://iseg2017.web.unc.edu). In this challenge, researchers were invited to propose and evaluate their methods to segment WM, GM, and CSF on 6-month infant brain MRI scans. Most of the top performing methods were based on deep learning techniques. In particular, deep learning methods based on CNNs have demonstrated outstanding performance in solving the segmentation of isointense infant brain MRI. Below, we review some of the commonly used CNN architectures used for such a segmentation task.

Convolutional neural networks (CNN): Unlike conventional deep neural networks where inputs are always in vector form, CNNs maintain and utilize the structural and spatial information among neighboring pixels or voxels in the input 2D or 3D images. Also, a CNN limits the degrees of freedom of the deep learning model through exploiting a local receptive field, weights sharing, and sub-sampling techniques (see Supplementary Fig. 1). Hence, this will result in models that can learn from limited datasets commonly present in medical applications in a way that is less prone to the problem of overfitting. Although CNNs were first introduced in 1989 [52], it did not gain interest until the introduction of deep CNN architectures, such as AlexNet [18] that accomplished remarkable results in the ImageNet [53] competition in 2012. AlexNet represents a currently typical CNN architecture that consists of a sequence of layers of convolution, pooling, activation, and fully connected (dense or classification) layers (see Supplementary Fig. 2). In the image classification task, AlexNet almost halved the error rates of the formerly best-performing methods [53]. Since then, CNN architectures are increasingly deeper, which resulted in improved error rates, often achieving human performance levels [19,54,55]. A convolutional layer is designed to detect localized features extracted at different locations in the input feature maps using learnable kernels or filters. The output of such convolutional layers is passed through non-linearities introduced by an activation function. A rectified linear unit (ReLU) and its alterations, such as Leaky ReLU, are among the most commonly used activation functions. These functions enabled the training of deeper models in a way that avoids vanishing gradient problems typically found when using the more traditional tanh and sigmoid activation functions. Pooling layers are used to downsample the feature maps using the maximum or the average of the predefined neighborhood as the value passed to the next layer. To classify an input image, the output scores of the final CNN layer are then related to a loss function (e.g., a cross-entropy loss that maps scores into a multinomial distribution over class labels). The network's parameters are learned by iteratively minimizing the loss function with parameters being updated (e.g., using stochastic gradient descent SGD) via back-propagation until convergence is reached.

Patch-based convolutional neural networks: As mentioned above, CNN based deep learning methods have accomplished great success in medical image segmentation, including isointense infant brain tissue segmentation.

The most straightforward approach employs a patch-based CNN architecture in which an NxN patch centered at each pixel is extracted from the input image and then used to train a CNN classification model to classify patches, represented by the center pixel, into the correct tissue class (i.e., WM, GM or CSF). Zhang et al. [56] first proposed using such patch-based CNN architecture to segment isointense stage infant brain images in which complementary and multi-modality information from T1w, T2w, and FA images are utilized as inputs to generate the final segmentation maps The input to the CNN architecture is the three input feature maps that correspond to the T1w, T2w, and FA 2D image patches (see Supplementary Fig. 3). It is then passed through three convolutional layers with increasing width before a fully connected layer is applied with a softmax activation function to obtain class probabilities of the center pixel in the input patch.

The patch size selection limits the performance of such patch-based approaches. In particular, patch size selection leads to a trade-off between localization and classification accuracy. Using large patches improves classification accuracy but leads to increasing localization errors caused by the additional pooling layers. On the other hand, utilizing smaller patches leads to better localization and lower classification performance due to the limited context information and increased sensitivity to noise. Additionally, these methods are hard to train due to the large number of parameters involved, and since CNNs need to be applied to a massive number of locations, such methods are very computationally expensive. The performance of patch-based architectures can be improved using a multi-scale approach, where multiple patch sizes are utilized to construct multiple pathways, and the output of these pathways is combined using a fully connected layer with a softmax activation function [40]. The larger scales keep implicit information about the voxel location in the scan, while the smaller scales provide detailed information about the local neighborhood of the voxel (see Supplementary Fig. 4). Note that the kernel size is adaptive in that larger kernels are used for the larger patches. Although such an approach addressed the patch size selection problem, the use of sliding windows results in regions being processed independently in a computationally inefficient way due to the many redundant convolutions and pooling operations.

Fully convolutional neural networks: To overcome the problems mentioned above, a fully convolutional neural network (FCN) architecture can be utilized. An FCN can be trained end-to-end with convolutional layers to produce dense pixel outputs with much less network parameter as it avoids the use of fully connected layers [57]. Specifically, the architecture of FCN is similar to that of a typical CNN with the last fully connected layer replaced by a convolutional layer with a large receptive field. Using FCN results in a significant reduction in the number of trained parameters and allows for learning representations and decisions based on local spatial context unlike the global context captured by the fully connected layers. Thus, FCNs can simplify and accelerate both the learning and the inference of segmentation deep neural networks. In particular, in this architecture type, a classification is given for each voxel of the input image similar to semantic segmentation. They include an encoder network that extracts compact high-level features using convolutional and pooling layers, and a decoder network that upsamples the higher-level features extracted by the encoder typically once using a fixed interpolation method [57]. This produces pixel-wise class probabilities that can be used to classify pixels and produce the required segmentation. Nie et al. [58] proposed multiple fully convolutional networks (mFCNs) to segment the isointense infant brain image using multimodal information of T1w, T2w, and FA images. Separate FCN networks were used for each modality, and instead of merely combining three modality data from the original low-level feature maps, a deep architecture was employed to effectively fuse their high-level representation from three modalities to produce the final segmentation (see Supplementary Fig. 5). Due to the imbalance of voxels assigned to each category, a weighted loss function was used giving weights that are inversely proportional to the fraction of voxels of the corresponding tissue class. Although such method avoided some of the drawbacks of the patch-based methods, its particular implementation relied on 2D slice data that does not account for the anatomical context in the slicing direction. Also, this method assumes that the high-level representations from different modalities provide complementary information, which is not always the case.

Dilated convolutions: To incorporate larger anatomical context, Moeskops and Pluim [59] explored the use of dilated convolutions for the segmentation of isointense stage infant brain MR images. Dilated CNNs are a particular type of CNN proposed for image segmentation that can realize a large receptive field while limiting the number of trainable parameters [60]. Dilated convolutions add another parameter to convolutional layers called the dilation rate (see Supplementary Fig. 6). Dilation rate controls the spacing between the kernel elements, and hence, for instance, a 3×3 kernel with a dilation rate of 2 will have a receptive field similar to that of a non-dilated 5×5 kernel, while only using nine parameters instead of 25. The proposed FCN network utilized a triplanar (axial, sagittal and coronal planes) dilated convolutions and 3D convolutions in four network branches, where the dilated branches use layers of 3×3 kernels with increasing dilation factors. All network branches used 2-channel input from the T1w and the T2w images.

U-Net: Zeng and Zheng [61] proposed an approach where a context guided, multi-stream, two-stage 3D FCN architecture is employed. As shown in Fig. 5, the first stage FCN-1 is used to learn tissue-specific probability maps from the input T1w and T2w images. After obtaining an initial segmentation, a distance map is computed for each brain tissue to model the spatial context information. The second stage FCN-2 is used to obtain the final segmentation based on the computed spatial contact information and the original multi-modality MR images. The design of both FCN-1 and FCN-2 are inspired by a famous FCN architecture that was developed for biomedical image segmentation called U-Net [62]. The U-Net modifies the original FCN architecture by utilizing both long and short skip connections similar to the ones introduced in residual networks. This enables the network to learn from fewer training images and produce more precise segmentation [19] (see Supplementary Fig. 7). Similar to U-Net, both FCN-1 and FCN-2 consist of an encoder part (contracting path) that learns compact feature representations and a decoder part (expanding path) that generates the segmentation results from the learned representations. The decoder part is applied to each modality separately, and high-level information from all modalities are fused at the beginning of the decoder path. To alleviate the potential gradient vanishing problem during training, which is more difficult for full 3D image data, two auxiliary branch classifiers were injected into the network in addition to the classifier of the main network. To solve the problem of having to train the network from scratch with limited annotated data, transfer learning ideas have been leveraged by pre-training the network on a related segmentation task.

Other CNN approaches: Dolz et al. [63] proposed the use of an ensemble of 3D CNN for the segmentation of isointense infant MR brain images. Encouraged by the recent success of dense networks [20], in which each layer is connected to every other layer in a feed-forward fashion, a semi-dense architecture was proposed to connect all convolutional layers directly to the network end as shown in Fig. 6. This semi-dense architecture allows efficient gradient flow during training, while also reducing the number of trainable parameters (one order of magnitude fewer parameters than a 3D U-Net architecture). The network investigated the use of parametric ReLU (PReLU) [64], which is a parametric version of a leaky ReLU that prevent saturation by learning an additional scaling coefficient. Bottleneck layers ($1 \times 1 \times 1$ convolutions) were used instead of fully connected layers to limit the number of parameters and ensure the fully convolutional nature of the network. Different strategies for combining information from multiple MR modalities, specifically, early fusion (Fig. 6(a)) and late fusion (Fig. 6(b)) strategies were investigated regarding improving network

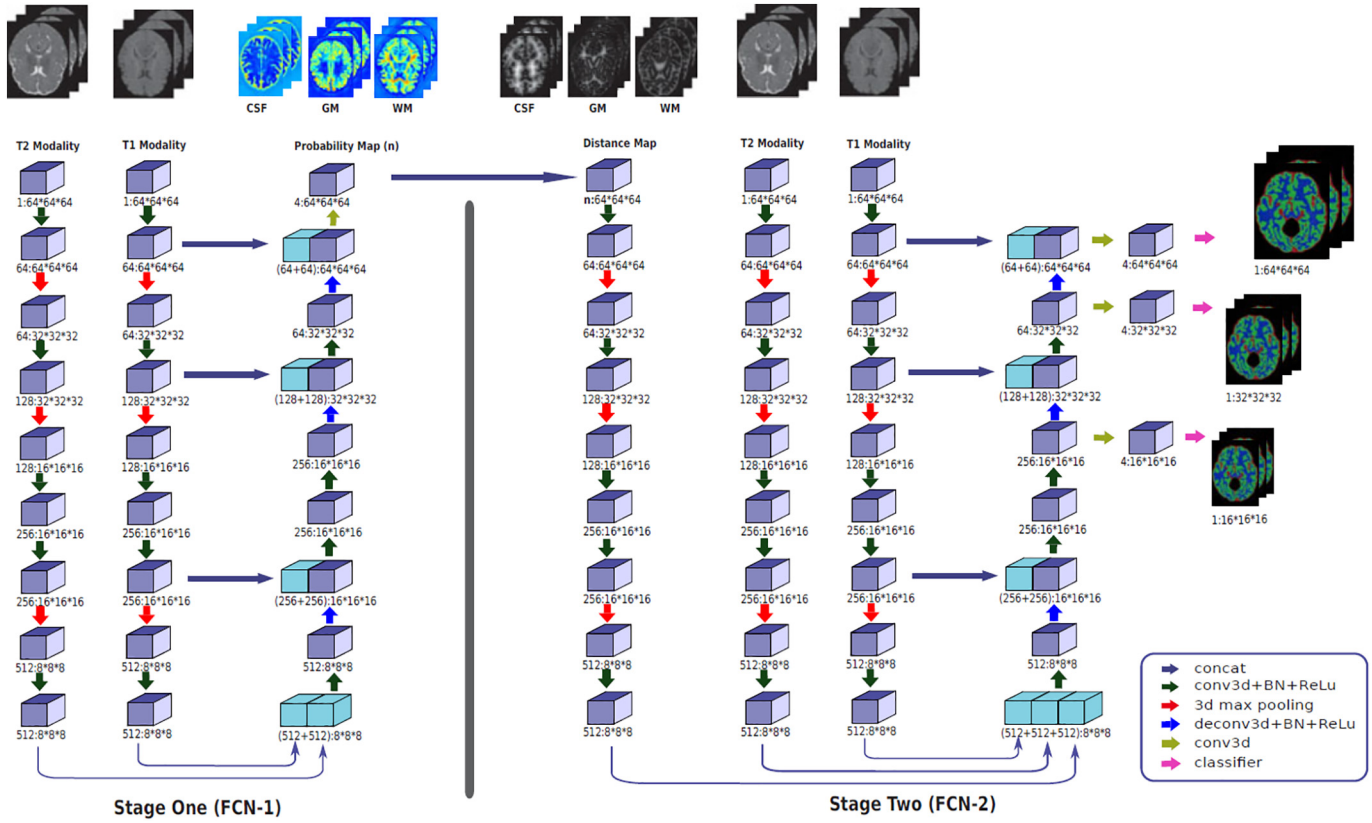


Fig. 5. Illustration of the two-stage 3D FCN architecture proposed by Zeng and Zheng [61] for the isointense infant brain segmentation in multi-modality MR images. The first stage FCN will produce an initial segmentation that in turn is used to model the spatial context map for each brain tissue using distance maps. The final segmentation is obtained in the second stage using both the spatial contact information and the original T1w and T2w images.
Image courtesy of [61].

robustness. The final segmentation was obtained by a majority voting of 10 of the previously described FCNs, which was shown to enhance the generalization of the proposed segmentation method. Ensemble predictors were employed to identify which images have uncertain segmentations, and moreover which brain regions need to be inspected and corrected manually by experts.

Finally, Nie et al. [65] extended their previous work [58] by extending the conventional FCN architectures to a full 3D framework. In the proposed 3D architecture, fewer pooling layers were used to mitigate the loss of localization information, while using more convolutional layers at the smallest convolution filters at $3 \times 3 \times 3$ to be able to capture details better. Similar to skip connections in the U-Net architecture, coarse feature maps were aggregated with dense feature maps produced by the deconvolution layers using an information pass-through (PT) mechanism which aims to model small structure tissues better. The designed PT operation involves the use of extra convolutional layers to fuse low-level and high-level features after the concatenation layers, which helps to produce more precise segmentation results. The authors also introduced an alternative mechanism to the PT operation that is referred to as convolution and concatenate (CC) subprocedure, in which the low-level representations are mapped to higher-level layers through a convolution operation followed by the concatenation of both levels of representation. These extra convolutional layers impose additional challenges during training, and hence batch normalization (BN) [66] layers are added to speed up the convergence of the network. An additional advantage of the proposed method is its ability to produce accurate segmentations with much faster testing speed compared to other competing 3D CNN architectures.

3. Deep learning for early disease prediction in infants

Early prediction of NDDs is one of the main goals of precision medicine applied to the developing, pediatric population. An early prediction of NDDs is expected to have a significant impact as it allows for early and more effective interventions, which in turn is expected to result in an improved prognosis. The development of sophisticated, noninvasive 3D medical imaging technologies such as MRI provides a way to extract biomarkers to assist in the early diagnosis of NDDs. Neuroanatomical abnormalities in the cerebral cortex are widely investigated by examining cortical brain morphometric measures such as cortical thickness and cortical surface area. Such cortical measures are usually extracted from finely-sampled cortical surfaces, forming a high-dimensional feature list with 100,000 or more features corresponding to the number of vertices in the cortical mesh (Fig. 7). Using such cortical features, several studies have shown that the brain undergoes a tremendous change during the first year of life [67,68]. For example, in studying ASD, It was demonstrated that infants who will later develop ASD show a variety of age-specific, brain differences between 6 and 24 months, as well as demonstrating early brain changes that correlate with ASD outcome and related behaviors measured at 24 months [69–73]. Although significant, these group-level differences and brain-behavior correlations do not allow for individual-level outcome prediction critical for application to clinical practice. It would be particularly beneficial if such extracted neuroimaging markers could predict the intervention that would be most beneficial to child's prognosis [74]. To achieve this goal, many studies have used traditional machine learning methods with MRI data to classify individuals with NDDs, particularly ASD or to understand their patterns of behavior [26,75–77]. However, there is a lack of studies that prospectively classify infants before the onset of the core features of the disorder due

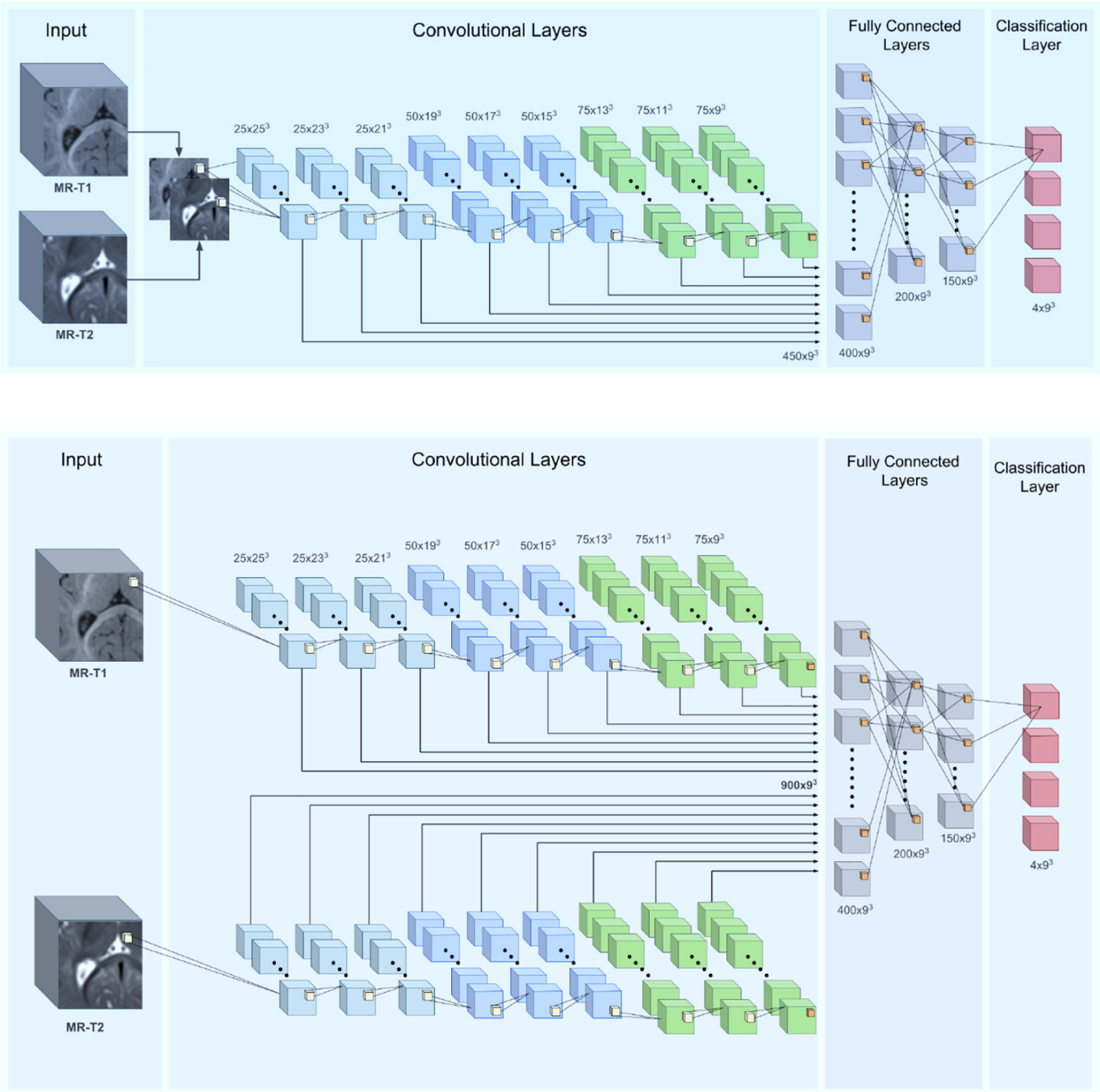


Fig. 6. An illustration of the ensemble of semi-dense 3D FCN proposed by Dolz et al. [63], where information from T1w and T2w images are fused using (a) early fusion strategy and (b) late fusion strategy.
Image courtesy of [63].

to a number of challenges when learning from MRI brain images; including (*1*) reliable extraction of relevant features from the cortical surface that are specific to the investigated disease, (*2*) determination of the optimal parsimonious subset of cortical features needed for an accurate disease diagnosis, (*3*) learning directly from a high-dimensional feature space in an end-to-end fashion without the need for a sub-optimal separate dimensionality reduction step, (*4*) design of classifiers that are able to utilize multi-modal biomarkers in predicting disease severity, and (*5*) identification of the regional brain regions and their interactions that contribute the most to the classifier performance.

There has been a limited effort to utilize deep learning techniques to tackle these challenges, which can be attributed to the lack of benchmarking infant MRI datasets. Indeed, there is a need for challenges sponsored by medical imaging conferences to encourage benchmarking and the development of new deep learning algorithms for early prediction of NDDs. Below, we will review recent efforts to apply deep learning techniques in the *early prediction of ASD*, which can be extended to serve as an application target for future benchmarking.

Can ASD be predicted from presymptomatic MRI data? A recent study showed that brain imaging in infants at high familial risk (HR) (by their having an older sibling with ASD) could predict a diagnosis of ASD with 80% positive predictive value (PPV) from earlier MRI data. This work utilized cortical surface features extracted from T1w and T2w brain images acquired at 6 and 12 months of age in order to

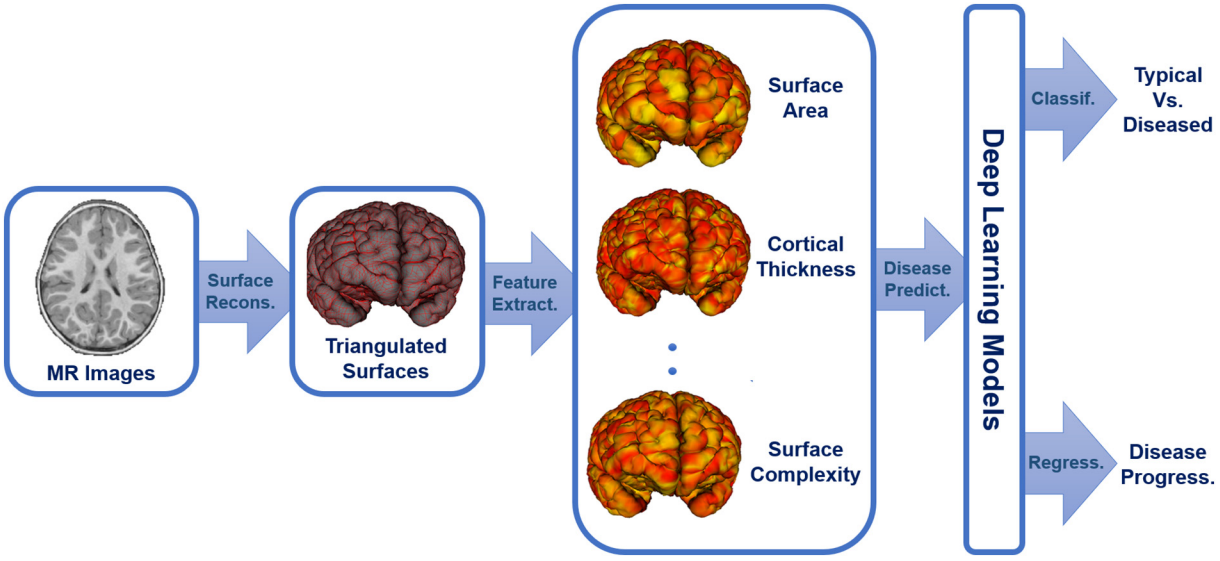


Fig. 7. An example deep learning-based framework for the early prediction of NDDs using high-dimensional cortical measurements extracted from infant MRI brain images.

accurately predicted ASD diagnosis at 24 months using a deep learning technique [11]. By comparison, behavioral features (even those measured as late as 18 months of age) have not produced individual-level diagnostic prediction that is of sufficient accuracy to be clinically useful [78].

In this prediction framework [11], brain tissue segmentation was computed using a standard expectation-maximization, atlas-moderated tissue classification [79] and cortical surfaces were then reconstructed from the 12-month MRI segmentations with an adapted version of the CIVET workflow [10]. The cortical surfaces consisted of 81,920 high-resolution triangle meshes (40,962 vertices) in each hemisphere, and cortical surface correspondence across-subjects was established via spherical registration to an average surface template [10]. The local surface area (SA) is computed at the middle surface, i.e., the surface that runs at the mid-distance between WM and outer GM surfaces to avoid over or under-representation of gyri or sulci [80]. The local cortical thickness (CT) is computed using the Euclidean distance between corresponding white and outer gray surface points [10]. Based on the extracted features, machine learning classifiers can be used to predict diagnostic outcomes. Such classifiers undergo a learning process in which they learn the correct category labels for each set of features. However, it is difficult to train stable classifiers using the high-dimensional feature space present in this study without overfitting.

Dimensionality reduction: One possible solution to overcome such an overfitting problem is to employ a supervised or unsupervised dimensionality reduction technique [81]. Unsupervised methods run the risk of losing relevant information while supervised methods tend to be more biased and therefore harder to generalize. Alternatively, if possible, dimensionality can be reduced by first summarizing the features via prior anatomical knowledge, such as summarizing vertex-wise cortical features via a cortical subdivision that divides the cortex into a mosaic of anatomically and/or functionally distinct, spatially adjacent areas using prior parcellation atlases [82]. The objective of dimensionality reduction is three-fold: improving the prediction performance of the predictors, providing faster and more cost-effective predictors, and providing a better understanding of the underlying process that generated the data.

In the framework proposed in [11], the dimensionality of CT and SA measurements were first summarized using the Brodmann-based automatic anatomic labeling atlas, namely AAL [83,84]. A deep learning model was then used as a second dimensionality reduction (Fig. 8(a)). The input feature vector was first binarized using a trained binary masking operation to allow performing an initial, unsupervised auto-encoder training procedure on individual two-layer greedy networks [85] (Fig. 8(b)). As shown in Fig. 8(c), a supervised training step is then applied to the autoencoder via a single classification layer with a single output representing the binary diagnosis label. The weights of the full network were fine-tuned using backpropagation treating the deep network as a traditional feed-forward neural network. After this supervised training step is completed, the last layer is removed, and the number of nodes in the last hidden layer represents the final dimension of the reduced dimension output. Finally, the reduced representation generated by the trained deep learning network along with the binary training labels is then used to train a two-class support vector machine (SVM) classifier.

Understanding the prediction: While an accurate prediction can be clinically relevant; it is also crucially important to provide information on how the prediction is made and how that generates a better understanding of the disease. To that end, identifying which input features defined in the high dimensional feature vector have the most significant impact on the generated lower dimension representation is highly relevant, yet a highly challenging problem for deep non-linear dimensionality reduction approaches. To solve this problem, the authors proposed an approach to rank the contribution of different features based on the learned weight matrices in the fully trained deep learning network. In particular, the proposed approach works backward over the deep learning network recognizing only those nodes in the previous layer (e.g., *l-1*) that denote > 50% of the weight contribution layer *l*, and this process is repeated till the input layer is reached.

Results: The recently reported results in Hazlett et al. [11] confirm that deep learning models can efficiently utilize morphological information from sMRI scans at 6 and 12 months to predict ASD diagnosis at 24 months at a clinically relevant level (88% sensitivity, 95% specificity, 81% positive predictive value and 97% negative predictive value). The proposed prediction pipeline outperformed other competing methods that utilized traditional machine learning methods to perform the dimensionality reduction step, including sparse learning [86] and principal component analysis [87]. Moreover, the proposed methods were shown to perform better than using an entirely deep learning architecture that combined both the dimensionality reduction and the classification stages in the proposed pipeline into one stage.

Improving the prediction: In a separate study on the same dataset, Shen et al. [73,88] showed that excessive volume of CSF in the subarachnoid space surrounding the cortical surface (i.e., extra-axial CSF

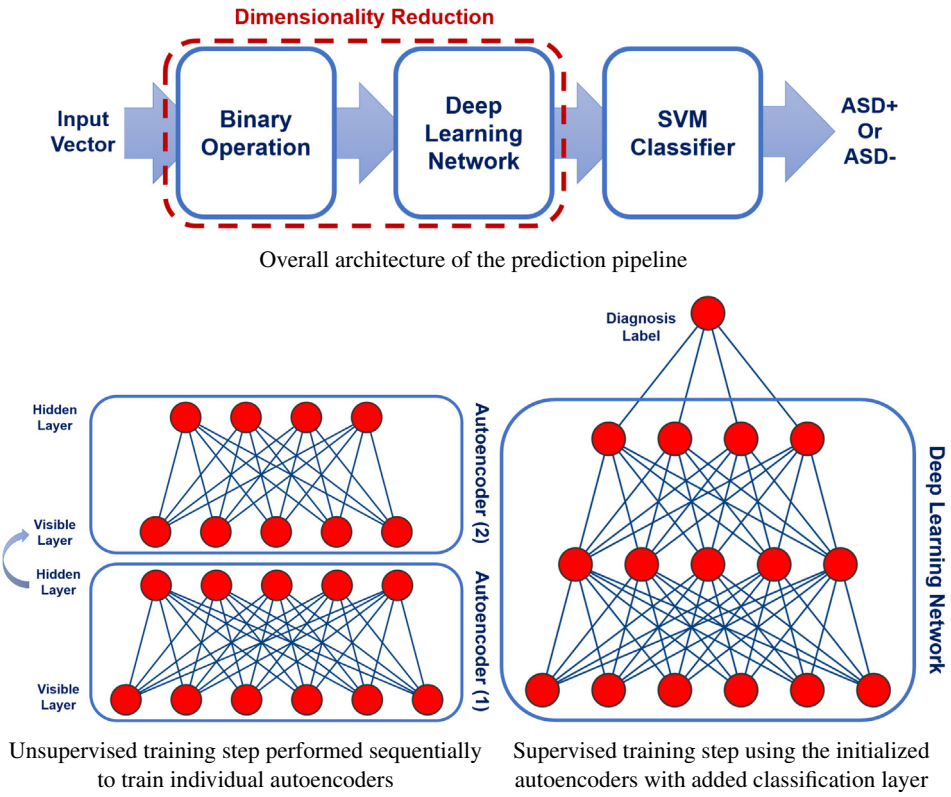


Fig. 8. The two-stage prediction pipeline proposed by Hazlett et al. [11] for early prediction of ASD.

or EA-CSF) is present in those infants who will later develop ASD already at 6-month of age. Collectively these two studies suggest that a substantially improved prediction performance is possible at an early age of 6-month by extending the approach in Hazlett et al. [11]. Directions to improve current prediction performance include (*1*) adding further cortical features, other than CT and SA, such as EA-CSF (*2*) employing alternative cortical parcellations other than AAL, (*3*) adapting a training scheme that accounts for the inherent sampling imbalance between the groups (kids at high familial risk for ASD have a 1 in 5 chance to develop ASD), and (*4*) extending the deep learning approach to include recent advanced techniques. Ongoing experiments are being conducted by our group to investigate all of these suggestions. Local measures of EA-CSF are thereby computed by summing probabilistic CSF volume along Laplacian streamlines [9]. Instead of summarizing local cortical features the geometrically based AAL atlas, the functionally-based atlas by Gordon et al. [89] is used. To account for class imbalance during the training phase, we utilize the oversampling technique such as SMOTE [90]. Finally, adding dropout and weight decay techniques can be used to regulate the network and overcome overfitting problems [91].

Convolutional networks on surfaces: Despite the high accuracy levels obtained in [11], one limitation of this work is its reliance on a prior cortical parcellation to reduce the dimensionality of the high dimensional cortical features, which is particularly problematic in infant studies as currently only adult parcellations are employed due to the scarcity of infant cortical surface parcellations. Hence, there is a need for data-driven classifiers that can directly learn from a high-dimensional feature space without needing a separate feature reduction step. One way to address this problem is to exploit CNNs ability to extract hierarchical abstractions directly from raw high-dimensional data with almost no prior knowledge and with fewer parameters. However, it is challenging to extend the use of CNNs to applications where the input features live in irregular non-Euclidean domains such as cortical brain surfaces. These challenges stem from the missing notion of a grid on a non-Euclidean surface, in addition to the need to adapt the convolutional filters locally while sliding across the input surface. Recently, Mostapha et al. [92] proposed a novel data-driven approach to extend classical CNNs to non-Euclidean manifolds such as cortical brain surfaces (see Fig. 9). The proposed CNN architecture introduced a general definition of surface kernels using a locally constructed localized grid, which in turn can be used to extract corresponding patches on the manifold. A local system of geodesic coordinates is utilized to establish a localized grid at each surface point in a way that adapts to the intrinsic shape of the underlying manifold. In addition to this newly developed surface convolution, novel surface subsampling and bottleneck layers are also introduced. To control the number of training parameters and to be able to learn higher-level representations, repeated surface subsampling is performed using adaptive edge collapsing that maintains an optimal approximation of the original geometry, local curvature, and geodesic distance information. Moreover, bottleneck layers are used to control the number of features maps to avoid the possibility of overfitting when training from a limited size dataset.

The usability of the proposed surface-CNN framework was confirmed in a separate study of Alzheimer's where significant differences in CT were expected. In particular, the proposed surface-CNN framework outperformed competing methods employing a variety of parcellation atlases or using traditional machine learning methods such as sparse coding to perform the required feature reduction before applying deep learning models.

Convolutions via 2D surface mapping: based Rather than applying convolutional networks directly on the surface, one can also define a mapping to bring the data into a space where conventional CNN methods can be applied. For example, closed surface data can be mapped onto a regular 2D image structure called a geometry image [93] (see Fig. 10). This allows the straightforward use of conventional image-based CNN architectures to learn directly from the high-dimensional input features without the complexities of the above-described surface CNN. However, experiments with geometry images of features

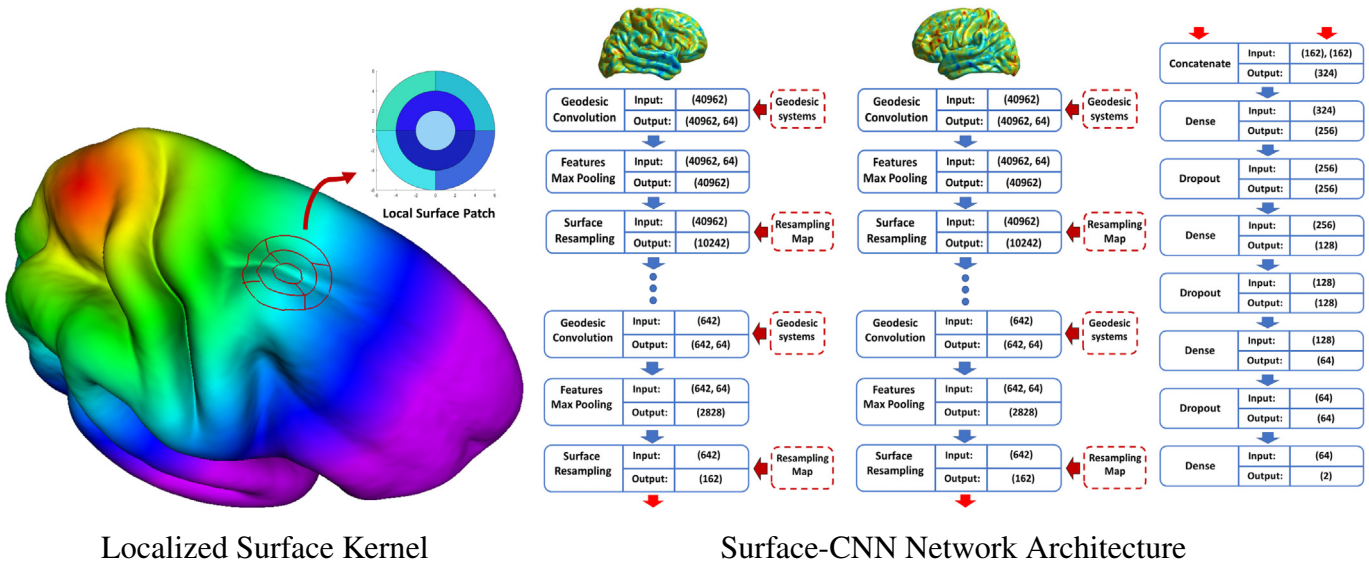


Fig. 9. CNN extension to cortical surfaces proposed by Mostapha et al. [92]. Several surface convolutional blocks are applied to each hemisphere in parallel followed by fully connected and dropout layers.

living on surfaces show that the performance level achieved by such a simplified convolutional approach is still significantly inferior to convolutional learning directly on the manifold [92]. This is most likely due to the non-isometric nature to the mapping into the 2D image space, that results in a surface specific distortion of the surface feature maps. As shown in Fig. 11, to demonstrate the generalizability of the proposed methods, the surface-CNN architecture was used to investigate early prediction of ASD using features extracted from 6-month subcortical brain surfaces (rather than cortical surfaces). Mainly, we examined the predictive ability of different morphological features such as local thickness and curvature information, which were derived from subcortical brain surfaces reconstructed using the SPHARM-PDM pipeline [7]. The obtained experimental results again confirm the superior performance of the surface-CNN framework compared to alternative methods that include the simplified geometry images CNN approach.

Dimensional outcome prediction: The above prediction focused on the diagnostic categorical outcome. The next step beyond diagnostic categories is to predict continuous dimensions of disease-related features. In particular, high risk (HR) infants who develop ASD will have a range of outcomes along multiple dimensions such as language, cognitive function, and joint attention. Also, approximately one-third of HR infants that do not develop ASD will still develop clinically-relevant cognitive and behavioral impairment. While these impairments are below a diagnostic cutoff, these infants may be particularly responsive to presymptomatic intervention. Indeed, current efforts are made by

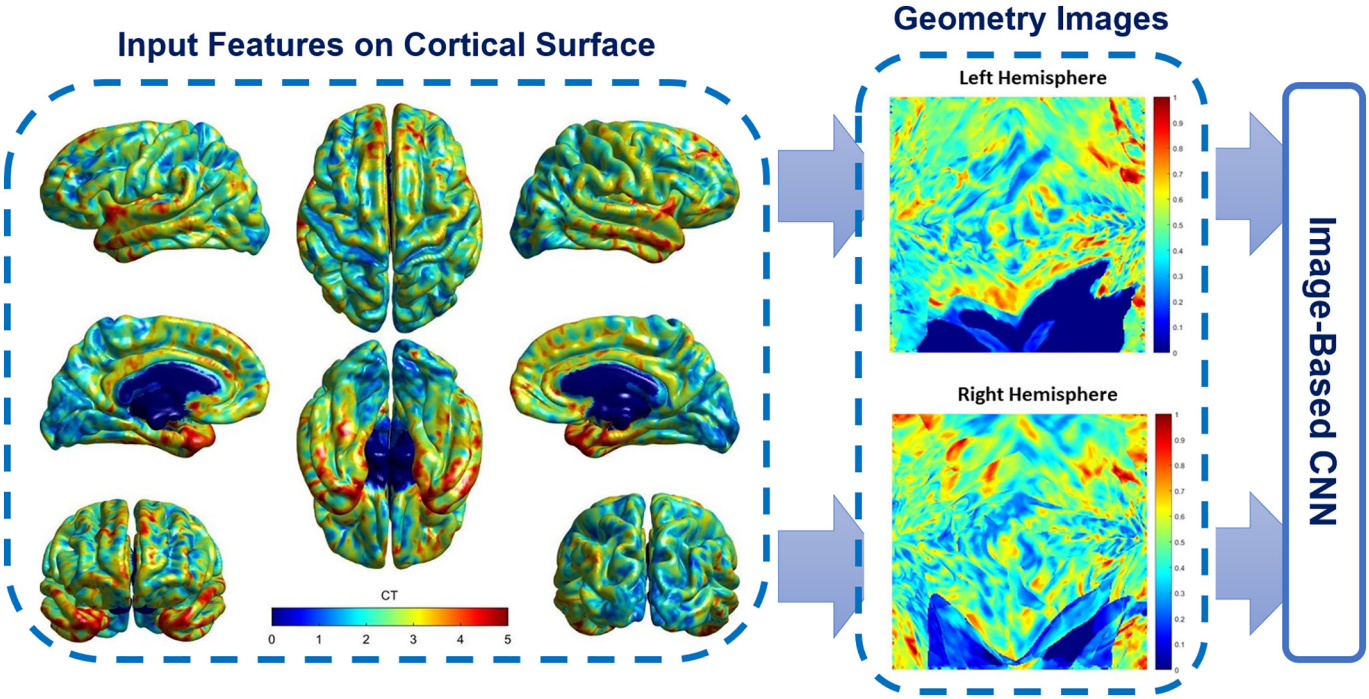


Fig. 10. An alternative to the CNN extension presented in Mostapha et al. [92], surfaces can be re-meshed onto a completely regular structure called a geometry image [93] which allow learning using conventional CNNarchitectures.

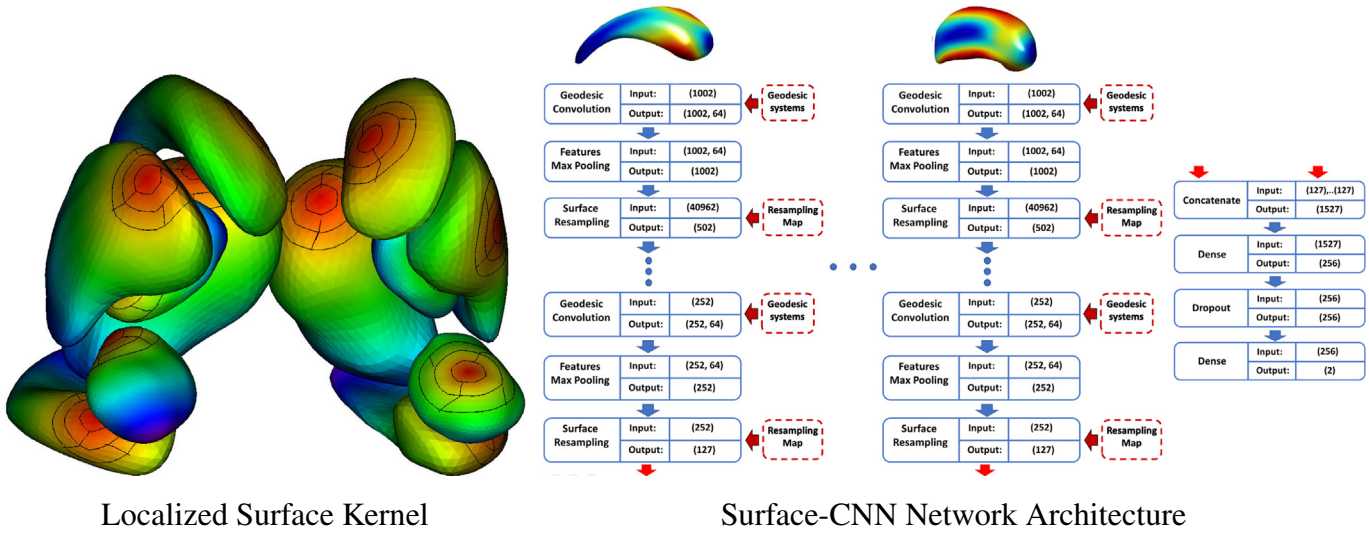


Fig. 11. The surface-CNN architecture proposed by Mostapha et al. [92] was applied for the early prediction of ASD using features extracted from 6-month subcortical brain surfaces.

our group to utilize MRI data of HR individuals in the first year of life to predict individual scores on later measures of cognition and behavior in several relevant domains. The capability of infant MRI brain imaging to predict later scores on dimensions of cognition and behavior, at an individual level, will facilitate the development and implementation of targeted, individualized, presymptomatic interventions in HR infants. Success with this goal would open up the possibility of testing personalized (versus one-size-fits-all) treatment approaches based on predicted individual profiles of social communication, language, motor, and cognitive ability. Evidence for the efficacy of presymptomatic interventions is currently scarce as no such interventions are effectively possible without an early predictive identification. Still, multiple randomized controlled trials indicate that early interventions such as joint attention-based interventions improve joint engagement and language outcomes in ASD [94,95]. If presymptomatic MRI predicts an individual scores on dimensional measures of a given intervention target (e.g., joint attention), treatment could target this domain of functioning more specifically.

In recent preliminary findings, traditional machine learning techniques such as support vector regression (SVR) [96] combined with an adapted (rank prediction) reliefF feature filtering [97] showed ability to predict standardized scores of language ability at 24-months using a combination of behavioral measures (MSEL [98] and AOSI [99]) and MRI-based measures including SA, CT, functional connectivity, and diffusion tractography in 6-month-old infants. However, there is still a need for deep learning approaches that can learn to predict such continues measures in a data-driven way without the need for the additional feature reduction step. In our ongoing research, we have preliminary findings that the deep learning identified ASD-related features to indirectly predict continuous scores of social ability at 24 months [100].

Multi-task learning for dimensional prediction: Encouraged by the above described results, we are currently investigating the use of multi-task learning (MTL) framework to directly predict continuous ASD-related scores using a combination of various cortical measurements. In addition to our ability to learn directly from high-dimensional feature space on cortical surfaces using our surface-CNN framework, MTL can leverage useful information contained in multiple related tasks of predicting different continuous scores to help improve the generalization performance of all the tasks, especially in the presence of limited datasets [101]. As shown in Fig. 12, in deep neural networks, MTL is typically done with either hard or soft parameter sharing of hidden layers. In hard parameter sharing, the hidden layers are shared between all tasks while several output layers are used for each task separately [102]. Hard sharing strategy is more commonly used since forcing the network to learn different tasks simultaneously will lead to relying on generalized representations which reduce the chance of overfitting on our original task [103]. On the other hand, soft parameter sharing allow each task to has its model and parameters, while regularization techniques (e.g., L_2 norm [104] or trace norm [105]) employed to keep the models parameters similar. Different multi-task deep learning methods have shown success in various medical image analysis application including image segmentation [106] and early prediction of degenerative diseases [107,108]. However, more research needs to be done to employ similar ideas in infant brain images effectively.

4. Open challenges in infant MRI

With recent development in the field of deep learning, many innovative methods have been proposed to improve infant MRI brain image processing and analysis as presented above. The success of deep learning is attributed to its ability to discover general morphological and textural features in a data-driven way that can handle different variabilities in infant MRI brain images that stem from complex brain anatomy and tissue appearance, imaging acquisition protocols, and pathological heterogeneity. In this section, we will discuss some of the open challenges to utilizing deep learning in MRI applications and discuss future directions to tackle these challenges.

4.1. Data size

In pediatric applications, available datasets are particularly small, as recruiting in such a vulnerable population is significantly more difficult than in adults or adolescents. This limits the ability of the deep learning methods to achieve their full power, especially when compared to their success in computer vision applications that utilize large-scale datasets such as ImageNet. In infant MRI brain imaging, the lack of publicly available datasets and high-quality labeled data imposes additional challenges that need to be addressed. Nevertheless, despite the small training datasets, deep learning employed for medical imaging data report reasonably satisfactory performance levels in numerous tasks [109]. To address the question of how much data is needed for training in medical image analysis, recent research indicates that the required dataset size really depends on the expected variability seen in the process being studied, and massive datasets are not always necessary [110]. Indeed, accurate classification can be accomplished

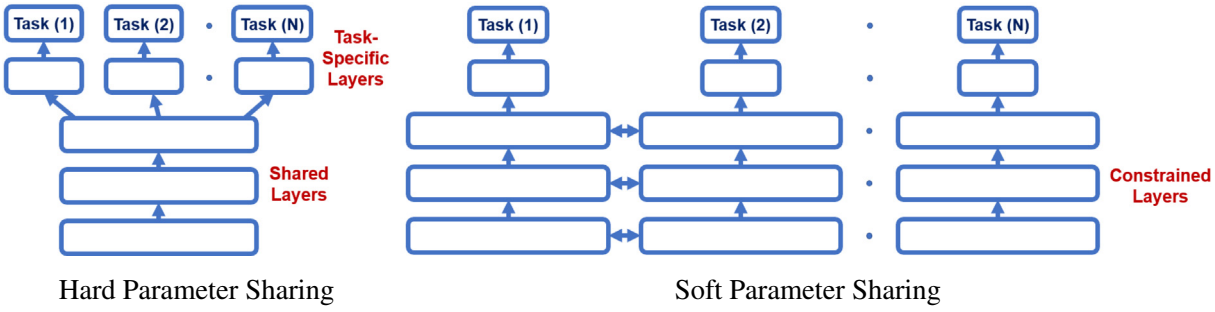


Fig. 12. The two most commonly used ways to perform multi-task learning (MTL) in deep neural networks.

with small medical imaging datasets due to the general image appearance homogeneity across different patients, contrasted with the near-infinite diversity of natural images (e.g., different breeds of dogs, colors, and camera poses) [111]. However, in the presence of NDDs in infant data, the heterogeneous nature of the MR images in the first year of life imposes challenges similar to that in natural images, and hence limited datasets have been a considerable obstacle in developing useful deep learning prediction models in the infant age range. Currently, there has been an effort to making data available through public research data repositories (e.g., NDAR for autism research), which will allow deep learning models to discover more generalized features in NDDs. Meanwhile, similar to computer vision tasks, attempts have been made to reduce the need for additional training data by carefully designing the deep learning framework (e.g., smaller filters for deeper layers) [54], different architecture groupings [112], or hyperparameter optimization [113].

To avoid the challenge of training deep learning models from scratch in the context of limited datasets, an alternative is to fine-tune a deep learning model that has been pre-trained using, for example, a large dataset of labeled natural images a technique referred to as transfer learning. In a recent study by Tajbakhsh et al. [114] showed that indeed that transferring knowledge from natural images to medical images is, in fact, possible, even with the relatively big difference between the source and target distributions. In their experiments on different applications and different imaging modalities, they demonstrated that fine-tuning deep CNNs are beneficial for medical image analysis, showing performance levels similar to fully trained CNNs and even outperforming fully trained networks in case of limited datasets. Moreover, the level of fine tuning (i.e., number of layers to be re-trained) was shown to be different from one application to another depending on similarities between applications and the amount of labeled data available for tuning. Nevertheless, it is still challenging to effectively utilize such 2D pre-trained networks for 3D infant MRI applications as such approach would ignore anatomical context in directions orthogonal to the 2D plane [115]. Recently, to solve this problem, Xu et al. [116] proposed to combine 3D-like contextual information by stacking successive 2D slices to form a multi-channel image, which is used as an input for an FCN pre-trained on ImageNet for natural image classification. Such approach showed success in providing a fast segmentation of both neonatal and adult brain 3D MR images, but renders the results sensitive to the head orientation of the subject in the MRI.

Alternatively, applying random transformations to the original data can effectively increase the dataset size a technique called data augmentation, which is a useful technique through simple transformations, such as flipping, rotation, translation, and cropping or even adding noise to the image [117]. Data augmentation aids in increasing the effective size of training samples and hence reduce overfitting by presenting random variations to the original data during training [18,118,119]. However, augmenting infant MRI datasets is still challenging due to the highly heterogeneous nature of the infant's brain, especially in the presence of neurodevelopmental disorders. Hence, there is still a need to develop data augmentation methods that are tailored to deal with additional challenges presented in infant data.

4.2. Class imbalance

Another challenge is the problem of the class imbalance in the medical applications [120]. Class imbalance refers to the number of training examples being skewed towards typical as compared to typical or pathological cases. This problem is exacerbated when developing diagnostic or predictive approaches for infants with NDDs as prevalence rates are commonly small in NDDs (e.g., in ASD $\leq 2\%$ for general population and $\leq 20\%$ for high-risk population). It has been recognized that the class imbalance problem has a substantial negative impact on training deep learning models. With imbalanced datasets, deep learning models tend to focus on learning the classes with a large number of examples, which leads to poor performance for the classes with a small number of examples. In a diagnostic setting based on medical image data, misclassification costs are typically unequal and classifying a diseased sample (minority class) as typical (majority class) implies significant consequences that should be avoided. Currently, there is no standardization or consensus on the effects of the class imbalance issues and how to mitigate this problem optimally, which could affect the reproducibility and accuracy of medical imaging research.

In classical machine learning, the class imbalance is a well-studied issue, and several methods have been proposed to solve this problem [120,121]. However, the methods used to address the class imbalance in case of traditional shallow models, are not always applicable to deep learning applications on complex medical application. Current literature shows a lack of studies addressing this issue for deep networks, as compared to classical machine learning algorithms. Current method for addressing class imbalance can be classified into three main types [122] (*1*) methods that operate on training set by altering its class distribution, (*2*) methods that work on the classifier or algorithmic level while keeping the training dataset unchanged, and (*3*) hybrid methods that combine the two previously described categories.

Oversampling: On the data level, oversampling methods are commonly used in deep learning, which in its basic form is called random minority oversampling in which samples from the minority classes are randomly replicated [123]. However, such simple oversampling method can lead to overfitting, and hence more advanced methods are needed, such as SMOTE or synthetic minority over-sampling technique [90]. In this method, artificial samples are created by interpolating neighboring data points, which has shown success in balancing training set in infant MRI applications [124–126]. Local synthetic instances (LSI) is an alternative method to SMOTE, which was proposed by Brown et al. [127] to augment high dimensional small sample size (HDSSS) infant MRI data to generate instances that are guaranteed to be near to real data instances. Recently, several extensions to SMOTE have been proposed, which may be promising to handle infant MRI datasets better. Such extensions include focusing on critical data points that are on the boundary between classes [128]. To decreases the between-class along with within-class imbalance, cluster-based oversampling methods will first cluster the samples in the dataset and then oversample each

cluster independently [129]. Other methods utilize boosting techniques to identify hard examples that in turn will be used to generate the required synthetic data [130]. Also, oversampling methods have been proposed for deep learning frameworks that ensure a uniform class distribution of each mini-batch [131].

Undersampling: Alternatively, undersampling can be used to solve the class imbalance problem by randomly removing samples from majority classes until a classes balance is reached [132]. However, in the context of limited datasets in infant MRI datasets, this technique is less popular as it discards a portion of available, highly valuable data. Some modifications have been proposed to overcome that by only removing redundant examples close to the boundary between classes [133]addressing or relabeling of some majority classes samples [134]. This can be achieved in deep learning models by using a weighted version of the loss function by introducing a balancing factor to impose the learning of minority classes samples [135]. Another modification of deep learning models to be cost sensitive is to adapt the learning rate in a way that is allowing examples that induce higher costs to contribute more to the update of weights [136]. The results obtained by this approach are similar to the oversampling described above.

One class learning: Another algorithmic strategy to achieve classes balance in challenging infant MRI datasets is what is referred to as one-class learning. In this strategy, the focus is to train models to recognize positive samples instead of discriminating between classes. This can be achieved using deep learning autoencoders that are trained to perform autoassociative mapping, and then new samples can be classified using a defined reconstruction error (e.g., squared sum of errors, Euclidean or Mahalanobis distance) [137,138].

Various hybrid approaches that combine multiple techniques from the previously described methods have been proposed to solve the class imbalance problem. For instance, SMOTEBoost is a method that combines boosting ideas with and SMOTE oversampling [139]. Recently, a technique that was introduced and successfully applied to CNN training for medical image segmentation task is two-phase training [140]. In this method, the network is first trained on a balanced dataset typically using an oversampling method, and then the output layers are fine-tuned on the original (imbalanced) data while keeping the same hyperparameters used in the first phase.

4.3. Interpretation

Deep learning algorithms have already exceeded human performance levels in many image recognition tasks, and it is probable that they will perform similarly in medical image analysis applications that involve infant MRI data. Such performance levels were achieved through highly flexible models with millions of weights that can learn an internal representation of the input data by optimizing a loss function. However, it is still challenging to compute an interpretation of how particular weights or inputs are contributing to the final model performance. Such interpretations are particularly crucial for successfully deploying deep learning models for early prediction of NDDs in a clinical setting.

In classification applications in the field of medical image analysis, features importance is typically determined and visualized by plotting the weights of a linear classifier [141] or plotting the *p*-values associated with these weights [142]. Such an approach ignores the connection to the input image and has been shown to lead to misleading interpretations and also not applicable to non-linear deep learning models [143].

Recently, methods for interpretation and understanding of deep learning models are increasingly proposed [144,145]. One way to provide insights into what deep learning models are via generating interpretable visualizations that capture the high-level concepts learned by the trained network. Two approaches have been proposed to achieve that, (*1*) finding input images that maximize a class score at the output layer in order to visualize how the network represents a specific class [146–148] or (*2*) visualizing feature maps that explain the network classification or decision in response to a particular input image. The second approach is better suited for models trained on infant MRI datasets and in particular in NDDs, which is are heterogeneous diseases in nature and hence require subject-specific interpretations.

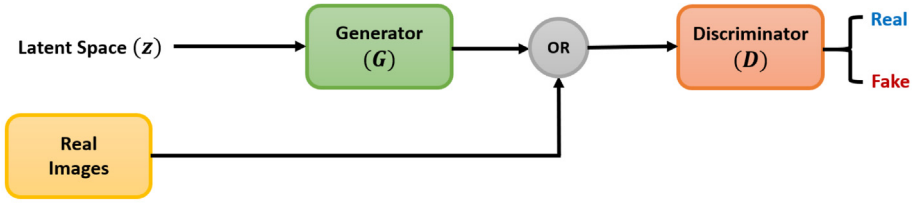
Saliency maps: One method to achieve this sample specific interpretation is via visualizing a class saliency map that is specific to a given image and class [147]. This is accomplished by determining how sensitive a class score is to a small perturbation in pixel values, which is computed by a single backpropagation pass to calculate the partial derivative of the class score with respect to the input pixels. The usefulness of such saliency maps was confirmed by employing them for weakly supervised object segmentation using classification CNNs. Another interpretation framework, called DeepLIFT, was proposed by Shrikumar et al. [149] in which importance scores are assigned based on explanations made based on differences between the output and some reference output regarding differences of the inputs from the corresponding reference inputs. This idea of using differences from a reference point permits information propagation even in the case of zero gradients, which is useful in case of deep networks with saturating activations like sigmoid or tanh. Another visualization method that was applied to MRI brain images was presented by Zintgraf et al. [150] in which conditional, multivariate modeling was used. In this method, the importance of input pixels was estimated via the correct class probability as a function of a patch occluding corresponding parts of the input image. Thus, the significance of a feature can be estimated based on how the output class probability changes when this feature is unknown or occluded. Such a method could be further improved by using more sophisticated generative models for conditional sampling, and also by incorporating spatial information (in infant MRI images for example) into the conditional distributions.

Layer-wise feature activity maps: Other class of methods try to understand individual decisions made by the classifier while assuming a black-box classifier or assuming a particular structure of how decisions are made [151–154]. A popular method introduced by Zeiler et al. [149] was explicitly designed for understanding CNNs through visualizing feature activity in intermediate layers. This is accomplished using a deconvolutional network described in [155], which has similar components to a CNN model (i.e., filtering and pooling operations) but in the reverse direction to map feature maps to corresponding input images. Unlike deconvolutional networks, no weights are learned and only used to probe the already trained CNN. Deconvolution operations are performed by using the transpose of the filters learned in the CNN model, while the non-invertible max-pooling operations are reversed through keeping a record of the locations of the maxima inside each pooling region in a set of variables also called switches. The obtained visualizations of intermediate feature layers were also shown to have a diagnostic role by helping to identify problems with the model and to come up with alternative architectures leading to better results. One limitation of this method is that it can only visualize a single activation at each layer, and there is still a need for methods that can visualize the joint activity present in a layer.

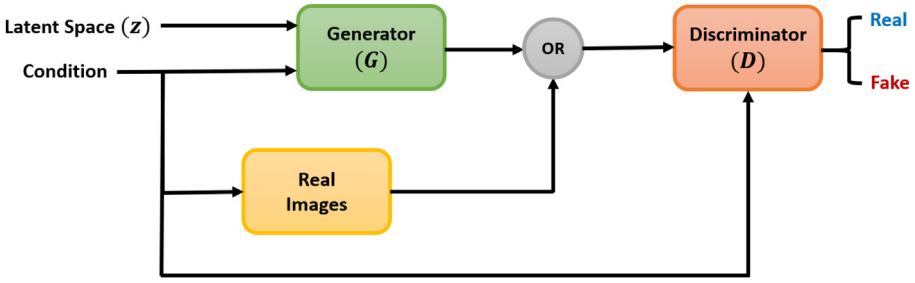
The interpretation methods described above can produce valuable insights into what infant predictive deep learning models have learned; however, there is little agreement on how these methods should be evaluated for benchmarking. One way is to evaluate interpretability in the context of a given application or using a proxy to provide a quantifiable evaluation [156]. Understanding how the model works is especially essential in medical applications where the underlying biological pathology of the studied disease is often still under research, as well as making the wrong decisions can be costly [157] as it may lead to incorrect or suboptimal subsequent interventions.

5. Towards generative models

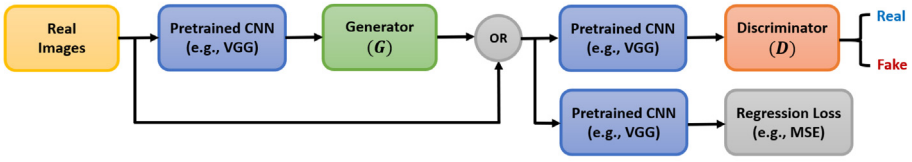
Although discriminative deep learning models have recently



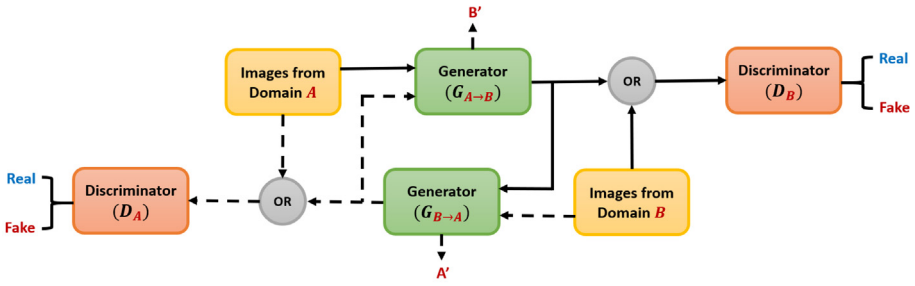
Deep Convolutional GAN (DCGAN) [168]



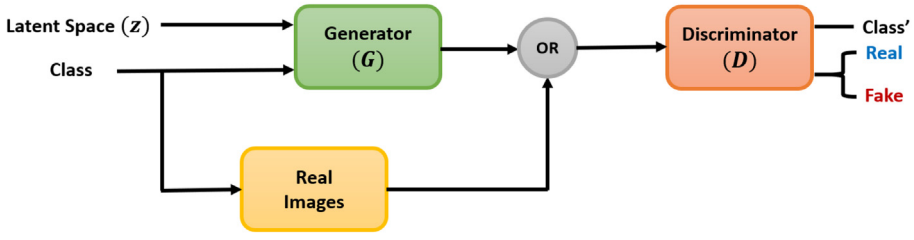
Conditional GAN (cGAN) [169]



Markovian GAN (MGAN) [170]



Cycle GAN (cycleGAN) [171]



Auxiliary Classifier GAN (AC-GAN) [172]

Fig. 13. Illustration of some examples of the generative adversarial networks (GAN) variants utilized for different tasks in medical image analysis.

reached near-human-level performance in a variety of tasks, the robustness of state-of-the-art deep classifiers is still an open question as it was shown that such models could be unstable to small perturbations in the input data [158]. Such lack in generalization performance can be attributed to the difference between discriminative models and generative models [159]. Discriminative models generalize well only when labeled data is plentiful [160], which is a critical limiting factor in the context of infant medical applications, where collecting vast amounts of annotated balanced training data is significant, rarely achieved challenge. In this context, generative modeling that can exploit unlabeled data in addition to labeled data can provide a relief to resolve this problem. However, traditional generative models have shown an inability to scale up to high dimensional datasets [161].

Recent advances in parameterizing generative models via deep learning models, combined with advancement in stochastic optimization techniques, have allowed scalable modeling of complex, high-dimensional data. Over the years, many revolutionary deep generative models have been proposed, which includes restricted Boltzmann machine (RBM) [162], deep Boltzmann machine (DBM) [163], deep belief networks [164], variational autoencoders (VAE) [165] and generative

adversarial networks (GAN) [166]. Regardless of what type of generative model, the aim is to learn the underlying, unknown true data distribution from the training set, which in turn is used to generate new data points with some variability. It is rarely possible to learn the exact distribution of the given data implicitly or explicitly, and thus instead, the goal is to model a distribution that is closely similar to the exact data distribution. To achieve that, the power of neural networks is leveraged in learning a function that can approximate the model distribution to the exact distribution. Deep generative models are promising to provide solutions for problems associated with limited and imbalanced datasets. Currently, the utilization of these methods to solve the challenges associated with learning from infant MRI datasets is still limited. Below, we will discuss current deep generative models and describe how they can be adapted to provide relief to the existing infant MRI datasets challenges.

Currently, VAE and GAN are the most commonly used architectures due to their ability to provide efficient and accurate models. VAEs explicitly try to approximate the true data distribution using a Bayesian inference encoder network and a decoder network, which provides a probabilistic version of an autoencoder. In particular, VAE adds a constraint to encoding the input data, namely that the encoded latent space variables are forced to approximate a normalized Gaussian distribution. The decoder network will then map this latent space into output data, and the combined network is trained by maximizing the lower bound of the data log-likelihood. One advantage of using VAEs is that there is a clear and recognized way to evaluate the quality of the model using the estimated log-likelihood. However, due to the strong assumptions and approximations involved, they may lead to suboptimal models, and the generated images tend to be more blurred than those coming from GANs [166].GANs and their extensions have shown to provide novel ways to tackle challenging medical image analysis problems such as medical image de-noising, reconstruction, segmentation, simulation, and classification. Furthermore, GANs ability to create synthetic images at extraordinary levels of realism promises to solve the scarcity problem of labeled data in the medical imaging field. Unlike VAE, GAN implicitly specifies probabilistic models that describe a stochastic procedure to directly generates data. Such a framework for constructing generative models can provide images that are sharp and compelling without having to specify a likelihood function. GANs contain two simultaneously-trained, competing models, which may be deep learning models such as CNNs. GANs training is based on a game theoretic scenario, where two players are competing in a zero-sum game. One network is a generator that generates synthetic images, while the other is called a discriminator with the task of determining if input images are real training images or if they are synthetic images from the generator. In this adversarial arrangement, the generator is trained by optimizing a minimax objective together with a discriminator, where it is desired at the end of training that the discriminator is no longer able to identify the difference between a real and a synthetically generated image.

An advantage of this GAN setting is that both generator and discriminator can be trained with backpropagation, without the need for unwieldy inference and Markov chains. Though GANs show such essential advantages over other generative models, training GANs is difficult as it may lead to oscillatory behavior and suffer from a problem referred to as mode collapse in which all latent space inputs are mapped to the same data point [167]. Due to vanishing gradients during training, GANs may also produce unstable models that generate different outputs for the same input. As shown in Fig. 13, other conditional variants along with a variety of extensions have been proposed to overcome the limitations of the vanilla GAN architecture, and have been successfully applied in the field of medical image analysis.

GAN variations: For example, the Deep Convolutional GAN (DCGAN) [168] architecture was proposed in which both the generator and discriminator utilize FCN architecture. DCGAN employ batch normalization and leaky ReLU to stabilize the performance of the network; however, it still does not address the mode collapse issue. In the conditional GAN (cGAN) architecture proposed by Mirza and Osindero [169], the generator is presented with random noise as well as some prior information, which was shown to improve training stability and quality of the generated output. Fast and high-quality style transfer, i.e., the task of recomposing images in the style of other images, can be accomplished using another conditional GAN architecture is called the Markovian GAN (MGAN) [170]. MGAN architecture utilize pre-trained CNN networks such as VGG19 to extract features to help to preserve the original image content while transferring the style excellently. To discover the underlying connection between two image domains, another GAN variant called cycleGAN was proposed by Zhu et al. [171] in which a cycle training algorithm is applied to unpaired data to extract key features of one image domain needed for a translation to another image domain. The auxiliary classifier GAN framework was introduced by Odena et al. [172] as an alternative to cGAN architecture, where the discriminator is tasked with reconstructing the prior information instead of feeding it to providing the generator and the discriminator networks. To overcome the mode collapse and instability problems when training GANs, the Wasserstein-GAN (WGAN) [173] architecture that uses the Earth Mover (ME) or Wasserstein-1 distance instead of the Jensen-Shannon (JS) divergence from the original GAN framework to compare the data distributions of generated and real images. Regardless of these theoretical advantages offered by WGAN, it is expected to lead to slow convergence in practical scenarios. The Least Squares GAN (LSGAN) [174] is another GAN variant that was proposed to solve the instability problem of GANs by adding some parameters to the loss function to overcome gradient vanishing.

GAN in medical image analysis: Various GAN-based architectures have been proposed to solve medical imaging problems [175,176]. For example, GANs were used to solve the scarcity of balanced labeled data in medical imaging by synthesizing medical images unconditionally or conditionally. Unconditional GANs have shown an ability to synthase realistically looking medical images, which can help with challenges such as the lack of labeled data, class imbalance, data augmentation [177] and data simulation [178]. For Example, DCGAN with vanilla training was shown to be able to learn to generate high-resolution 2D T1-w MRI brain images from even a small number of samples [179]. On the other hand, conditional GAN frameworks were utilized to perform image synthesis by conditioning both on prior knowledge such as metadata, rather than noise alone. In Nie et al. citenie2017medical, CT brain images were synthesized from corresponding MR brain images using a network that combined both a cascade of 3D FCN with a GAN and to remove the need for paired training data. Wolterink et al. [180] proposed to use cycleGAN architecture to achieve the same. Recently, Yang et al. [181] proposed a 2D MRI cross-modality generation framework that leverages a cGAN architecture, which can cope with various MRI prediction tasks. The proposed generative framework was shown to provide an efficient alternative to previously developed discriminative based deep learning MRI translation methods [182,183].

A GAN architecture can also provide a viable alternative for current segmentation techniques utilizing U-Net architecture for the generator and a combined multi-task loss function. Xue et al. proposed a U-Net GAN-based framework called SegAN [184] in which pixel dependencies are learned using a multi-scale loss function. Also, GANs training strategy was adopted by Moeskops et al. [185] to enhance the performance of their FCN segmentation approach, and similarly Zhao et al. [186] showed that synthetic images information could enhance the segmentation of the bony structure in brain MRI images using a proposed architecture called Deep-supGAN. Anomaly detection is another area of medical imaging where GANs can be leveraged to overcome the need for a significant amount of labeled training data. In particular, GANs can be first trained on normal images, and then applied on diseased images leading to incomplete reconstructions that can be used to recover image regions of an anomaly. An example of that is given in. [187–189] in which an unsupervised GAN-based architecture called

AnoGAN [190] was adapted and applied to MRI brain images to discover brain regions associated with the investigated disease which in turn can be used for classification purposes. Finally, GANs can also be used to identify images with lack of adequate image quality or coverage as a way to automatically identify unuseable images [191]. Though recently GANs been effectively utilized for various tasks in medical image analysis, there is still a lack of studies adapting and applying such techniques in infant data either in the original Euclidian image domain or in non-Euclidian feature domains. Given the potential of GANs, in our research, we work towards precisely that.

6. Conclusions

While deep learning based methods have made significant strides in medical imaging applications, there are still some open problems, and relatively few methods have been applied in infant MRI data. MR images at infancy display many challenges due to the inhomogeneous tissue appearance across the image, the considerable variability of image appearance in scans from neonates up to 1-year-old subjects, as well as the low signal to noise setting. On the one hand, these challenges may explain the relative scarcity of publications. On the other hand, these challenges are difficult to address with non-deep learning methods, and the potential of deep learning likely allows researchers to overcome them. The reward for such success, particularly concerning applications to predictive analysis in neurodevelopmental disorders, is significantly high and is expected to allow for presymptomatic treatment interventions that could significantly alleviate disease symptoms later on.

In this document, we presented the current success in applying deep learning approaches in the infant MRI setting in two example applications, infant brain tissue segmentation at the isointense stage and presymptomatic disease prediction in ASD. Both applications are excellent examples where traditional approaches of image analysis and machine learning are not strong enough, but deep learning approaches are highly successful due to their ability to learn non-linear, complex relationships in variable, heterogeneous input data.

Open challenges exist such as low data size restrictions, class imbalance problems, and lack of interpretability of the resulting deep learning solutions. We discussed how existing solutions could be adapted to approach these issues as well as how generative models in particular seem to be a particularly strong contender to address the first two of these challenges. These solutions have though yet to be significantly applied in infant MRI data. Such research is ongoing in our lab as well as other labs around the world.

Acknowledgement

Funding was provided by the IBIS (Infant Brain Imaging Study) Network, an NIH funded Autism Center of Excellence (HDO55741) that consists of a consortium of 7 Universities in the U.S. and Canada, the following NIH grants U54HDO79124, R01EB021391, and K12-HD001441, U01 AG024904, as well as the SALT (Shape Analysis Toolbox for Medical Image Computing Projects) NIH grant (R01EB021391).

References

[1] A. P. Association, et al., American psychiatric association dsm-5 development, Proposed revisions/somatic symptom disorders/J 2.
[2] Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM. Fsl. Neuroimage 2012;62(2):782–90.
[3] Dai Y, Shi F, Wang L, Wu G, Shen D. ibeat: a toolbox for infant brain magnetic resonance image processing. Neuroinformatics 2013;11(2):211–25.
[4] Penny WD, Friston KJ, Ashburner JT, Kiebel SJ, Nichols TE. Statistical parametric mapping: The analysis of functional brain images. Elsevier; 2011.
[5] Fischl B. Freesurfer. Neuroimage 2012;62(2):774–81.
[6] Ad-Dabbagh Y, Lyttelton O, Muehlboeck J, Lepage C, Einarson D, Mok K, et al. The civet image-processing environment: A fully automated comprehensive pipeline for anatomical neuroimaging research. Proceedings of the 12th annual meeting of the organization for human brain mapping. vol. 2266. 2006. Florence, Italy.
[7] Styner M, Oguz I, Xu S, Brechbühler C, Pantazis D, Levitt JJ, et al. Framework for the statistical shape analysis of brain structures using spharm-pdm. Insight J. 2006;1071:242.
[8] Kim SH, Lyu I, Fonov VS, Vachet C, Hazlett HC, Smith RG, et al. Development of cortical shape in the human brain from 6 to 24 months of age via a novel measure of shape complexity. NeuroImage 2016;135:163–76.
[9] Mostapha M, Shen MD, Kim S, Swanson M, Collins DL, Fonov V, et al. A novel framework for the local extraction of extra-axial cerebrospinal fluid from mr brain images. Medical imaging 2018: Image processing. vol. 10574. International Society for Optics and Photonics; 2018. p. 105740V.
[10] Kim JS, Singh V, Lee JK, Lerch J, Ad-Dab'bagh Y, MacDonald D, et al. Automated 3-d extraction and evaluation of the inner and outer cortical surfaces using a Laplacian map and partial volume effect classification. Neuroimage 2005;27(1):210–21.
[11] Hazlett HC, Gu H, Munsell BC, Kim SH, Styner M, Wolff JJ, et al. Early brain development in infants at high risk for autism spectrum disorder. Nature 2017;542(7641):348.
[12] B. R. Howell, M. A. Styner, W. Gao, P.-T. Yap, L. Wang, K. Baluyot, E. Yacoub, G. Chen, T. Potts, A. Salzwedel, et al., The UNC/UMN baby connectome project (bcp): An overview of the study design and protocol development, NeuroImage.
[13] Wernick MN, Yang Y, Brankov JG, Yourganov G, Strother SC. Machine learning in medical imaging. IEEE Signal Process. Mag. 2010;27(4):25–38.
[14] Wang S, Summers RM. Machine learning and radiology. Med. Image Anal. 2012;16(5):933–51.
[15] Deo RC. Machine learning in medicine. Circulation 2015;132(20):1920–30.
[16] LeCun Y, Bengio Y, Hinton G. Deep learning. nature 2015;521(7553):436.
[17] Schmidhuber J. Deep learning in neural networks: an overview. Neural Netw. 2015;61:85–117.
[18] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems. 2012. p. 1097–105.
[19] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 770–8.
[20] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. CVPReditor. Densely connected convolutional networks. vol. 1. 2017. p. 3.
[21] Akkus Z, Galimzianova A, Hoogi A, Rubin DL, Erickson BJ. Deep learning for brain MRI segmentation: state of the art and future directions. J. Digit. Imaging 2017;30(4):449–59.
[22] Yang X, Kwitt R, Styner M, Niethammer M. Quicksilver: fast predictive image registration–a deep learning approach. NeuroImage 2017;158:378–96.
[23] Suk H-I, Lee S-W, Shen D, Initiative ADN, et al. Hierarchical feature representation and multimodal fusion with deep learning for ad/mci diagnosis. NeuroImage 2014;101:569–82.
[24] Kooi T, Litjens G, van Ginneken B, Gubern-Mérida A, Sánchez CI, Mann R, et al. Large scale deep learning for computer aided detection of mammographic lesions. Med. Image Anal. 2017;35:303–12.
[25] Hoo-Chang S, Roth HR, Gao M, Lu L, Xu Z, Nogues I, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans. Med. Imaging 2016;35(5):1285.
[26] Mostapha M, Casanova MF, Gimel¢farb G, El-Baz A. Towards non-invasive image-based early diagnosis of autism. International conference on medical image computing and computer-assisted intervention. Springer; 2015. p. 160–8.
[27] G. Li, L. Wang, P.-T. Yap, F. Wang, Z. Wu, Y. Meng, P. Dong, J. Kim, F. Shi, I. Rekik, et al., Computational neuroanatomy of baby brains: a review, Neuroimage.
[28] Hazlett HC, Gu H, McKinstry RC, Shaw DW, Botteron KN, Dager SR, et al. Brain volume findings in 6-month-old infants at high familial risk for autism. Am. J. Psychiatry 2012;169(6):601–8.
[29] Paus T, Collins D, Evans A, Leonard G, Pike B, Zijdenbos A. Maturation of white matter in the human brain: a review of magnetic resonance studies. Brain Res. Bull. 2001;54(3):255–66.
[30] Makropoulos A, Counsell SJ, Rueckert D. A review on automatic fetal and neonatal brain mri segmentation. NeuroImage 2018;170:231–48.
[31] Makropoulos A, Ledig C, Aljabar P, Serag A, Hajnal JV, Edwards AD, et al. Automatic tissue and structural segmentation of neonatal brain mri using expectation-maximization, MICCAI Grand Chall. Neonat. Brain Segment. 2012;2012:9–15.
[32] Beare RJ, Chen J, Kelly CE, Alexopoulos D, Smyser CD, Rogers CE, et al. Neonatal brain tissue classification with morphological adaptation and unified segmentation. Front. Neuroinform. 2016;10:12.
[33] Liu M, Kitsch A, Miller S, Chau V, Poskitt K, Rousseau F, et al. Patch-based augmentation of expectation–maximization for brain MRI tissue segmentation at arbitrary age after premature birth. NeuroImage 2016;127:387–408.
[34] Moeskops P, Benders MJ, ChitÇŽ SM, Kersbergen KJ, Groenendaal F, de Vries LS, et al. Automatic segmentation of mr brain images of preterm infants using supervised classification. NeuroImage 2015;118:628–41.
[35] Sanroma G, Benkarim OM, Piella G, Ballester MÁG. Building an ensemble of complementary segmentation methods by exploiting probabilistic estimates. International workshop on machine learning in medical imaging. Springer; 2016. p. 27–35.
[36] Weisenfeld NI, Warfield SK. Automatic segmentation of newborn brain MRI. Neuroimage 2009;47(2):564–72.
[37] Kim H, Lepage C, Evans AC, Barkovich AJ, Xu D. Neocivet: Extraction of cortical surface and analysis of neonatal gyrification using a modified civet pipeline.

International conference on medical image computing and computer-assisted intervention. Springer; 2015. p. 571–9.
[38] Wang L, Shi F, Yap P-T, Gilmore JH, Lin W, Shen D. 4d multi-modality tissue segmentation of serial infant images. PLoS One 2012;7(9):e44596.
[39] Wang L, Shi F, Li G, Gao Y, Lin W, Gilmore JH, et al. Segmentation of neonatal brain mr images using patch-driven level sets. NeuroImage 2014;84:141–58.
[40] Moeskops P, Viergever MA, Mendrik AM, de Vries LS, Benders MJ, Išgum I. Automatic segmentation of mr brain images with a convolutional neural network. IEEE Trans. Med. Imaging 2016;35(5):1252–61.
[41] Rajchl M, Lee MC, Oktay O, Kamnitsas K, Passerat-Palmbach J, Bai W, et al. Deepcut: object segmentation from bounding box annotations using convolutional neural networks. IEEE Trans. Med. Imaging 2017;36(2):674–83.
[42] Wang L, Shi F, Gao Y, Li G, Gilmore JH, Lin W, et al. Integration of sparse multi-modality representation and anatomical constraint for isointense infant brain mr image segmentation. NeuroImage 2014;89:152–64.
[43] Mostapha M, Alansary A, Soliman A, Khalifa F, Nitzken M, Khodeir R, et al. Atlas-based approach for the segmentation of infant dti mr brain images. Biomedical imaging (ISBI), 2014 IEEE 11th international symposium on. IEEE; 2014. p. 1255–8.
[44] Mostapha M, Soliman A, Khalifa F, Elnakib A, Alansary A, Nitzken M, et al. A statistical framework for the classification of infant dt images. In: Image Processing (ICIP)editor. 2014 IEEE international conference on. IEEE; 2014. p. 2222–6.
[45] Ismail M, Mostapha M, Soliman A, Nitzken M, Khalifa F, Elnakib A, et al. Segmentation of infant brain mr images based on adaptive shape prior and higher-order MGRF. Image processing (ICIP), 2015 IEEE international conference on. IEEE; 2015. p. 4327–31.
[46] Dolz J, Massoptier L, Vermandel M. Segmentation algorithms of subcortical brain structures on mri for radiotherapy and radiosurgery: a survey. IRBM 2015;36(4):200–12.
[47] Xue H, Srinivasan L, Jiang S, Rutherford M, Edwards AD, Rueckert D, et al. Automatic segmentation and reconstruction of the cortex from neonatal MRI. Neuroimage 2007;38(3):461–77.
[48] Melbourne A, Cardoso MJ, Kendall GS, Robertson NJ, Marlow N, Ourselin S. Neobrains12 challenge: Adaptive neonatal mri brain segmentation with myelinated white matter class and automated extraction of ventricles i–iv. MICCAI grand challenge: Neonatal brain segmentation (NeoBrainSI2). 2012. p. 16–21.
[49] Wang L, Shi F, Lin W, Gilmore JH, Shen D. Automatic segmentation of neonatal images using convex optimization and coupled level sets. NeuroImage 2011;58(3):805–17.
[50] Vardhan A, Fishbaugh J, Vachet C, Gerig G. Longitudinal modeling of multi-modal image contrast reveals patterns of early brain growth. International conference on medical image computing and computer-assisted intervention. Springer; 2017. p. 75–83.
[51] Wang L, Li G, Adeli E, Liu M, Wu Z, Meng Y, et al. Anatomy-guided joint tissue segmentation and topological correction for 6-month infant brain MRI with risk of autism. Hum. Brain Mapp. 2018;39(6):2609–23.
[52] LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, et al. Backpropagation applied to handwritten zip code recognition. Neural Comput. 1989;1(4):541–51.
[53] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. Imagenet large scale visual recognition challenge. Int. J. Comput. Vis. 2015;115(3):211–52.
[54] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. p. 1–9.
[55] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, (arXiv preprint arXiv:1709.01507 7)
[56] Zhang W, Li R, Deng H, Wang L, Lin W, Ji S, et al. Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. NeuroImage 2015;108:214–24.
[57] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. p. 3431–40.
[58] Nie D, Wang L, Gao Y, Sken D. Fully convolutional networks for multi-modality isointense infant brain image segmentation. Biomedical imaging (ISBI), 2016 IEEE 13th international symposium on. IEEE; 2016. p. 1342–5.
[59] P. Moeskops, J. P. Pluim, Isointense Infant Brain MRI Segmentation With a Dilated Convolutional Neural Network, (arXiv preprint arXiv:1708.02757)
[60] F. Yu, V. Koltun, Multi-scale context aggregation by dilated convolutions, (arXiv preprint arXiv:1511.07122).
[61] Zeng G, Zheng G. Multi-stream 3d FCN with multi-scale deep supervision for multi-modality isointense infant brain mr image segmentation. Biomedical imaging (ISBI 2018), 2018 IEEE 15th international symposium on. IEEE; 2018. p. 136–40.
[62] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. International conference on medical image computing and computer-assisted intervention. Springer; 2015. p. 234–41.
[63] J. Dolz, C. Desrosiers, L. Wang, J. Yuan, D. Shen, I. B. Ayed, Deep cnn ensembles and suggestive annotations for infant brain mri segmentation, arXiv preprint arXiv:1712.05319.
[64] He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. Proceedings of the IEEE international conference on computer vision. 2015. p. 1026–34.
[65] D. Nie, L. Wang, E. Adeli, C. Lao, W. Lin, D. Shen, 3-d fully convolutional networks for multimodal isointense infant brain image segmentation, IEEE transactions on cybernetics.
[66] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, arXiv preprint arXiv:1502.03167.
[67] Knickmeyer RC, Gouttard S, Kang C, Evans D, Wilber K, Smith JK, et al. A structural MRI study of human brain development from birth to 2 years. J. Neurosci. 2008;28(47):12176–82.
[68] Deoni SC, Mercure E, Blasi A, Gasston D, Thomson A, Johnson M, et al. Mapping infant brain myelination with magnetic resonance imaging. J. Neurosci. 2011;31(2):784–91.
[69] Wolff JJ, Gu H, Gerig G, Elison JT, Styner M, Gouttard S, et al. Differences in white matter fiber tract development present from 6 to 24 months in infants with autism. Am. J. Psychiatry 2012;169(6):589–600.
[70] Elison JT, Paterson SJ, Wolff JJ, Reznick JS, Sasson NJ, Gu H, et al. White matter microstructure and atypical visual orienting in 7-month-olds at risk for autism. Am. J. Psychiatry 2013;170(8):899–908.
[71] Lewis JD, Evans A, Pruett J, Botteron K, Zwaigenbaum L, Estes A, et al. Network inefficiencies in autism spectrum disorder at 24 months. Transl. Psychiatry 2014;4(5):e388.
[72] Wolff JJ, Gerig G, Lewis JD, Soda T, Styner MA, Vachet C, et al. Altered corpus callosum morphology associated with autism over the first 2 years of life. Brain 2015;138(7):2046–58.
[73] Shen MD, Kim SH, McKinstry RC, Gu H, Hazlett HC, Nordahl CW, et al. Increased extra-axial cerebrospinal fluid in high-risk infants who later develop autism. Biol. Psychiatry 2017;82(3):186–93.
[74] Gabrieli JD. Dyslexia: a new synergy between education and cognitive neuroscience. science 2009;325(5938):280–3.
[75] Katuwal GJ, Baum SA, Michael AM. Early brain imaging can predict autism: application of machine learning to a clinical imaging archive. bioRxiv 2018:1–16. 471169.
[76] Peng X, Lin P, Zhang T, Wang J. Extreme learning machine-based classification of ADHD using brain structural MRI data. PLoS One 2013;8(11):e79476.
[77] J.-W. Kim, V. Sharma, N. D. Ryan, Predicting methylphenidate response in ADHD using machine learning approaches, Int. J. Neuropsychopharmacol. 18 (11).
[78] Chawarska K, Shic F, Macari S, Campbell DJ, Brian J, Landa R, et al. 18-month predictors of later outcomes in younger siblings of children with autism spectrum disorder: a baby siblings research consortium study. J. Am. Acad. Child Adolesc. Psychiatry 2014;53(12):1317–27.
[79] J. Wang, C. Vachet, A. Rumple, S. Gouttard, C. Ouziel, E. Perrot, G. Du, X. Huang, G. Gerig, M. Styner, Multi-atlas segmentation of subcortical brain structures via the autoseg software pipeline, Front. Neuroinform. 8.
[80] Van Essen DC. A population-average, landmark-and surface-based (pals) atlas of human cerebral cortex. Neuroimage 2005;28(3):635–62.
[81] Van Der Maaten L, Postma E, Van den Herik J. Dimensionality reduction: a comparative. J. Mach. Learn. Res. 2009;10:66–71.
[82] Glasser MF, Coalson TS, Robinson EC, Hacker CD, Harwell J, Yacoub E, et al. A multi-modal parcellation of human cerebral cortex. Nature 2016;536(7615):171–8.
[83] Achard S, Salvador R, Whitcher B, Suckling J, Bullmore E. A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. J. Neurosci. 2006;26(1):63–72.
[84] Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 2002;15(1):273–89.
[85] Lee H, Grosse R, Ranganath R, Ng AY. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. Proceedings of the 26th annual international conference on machine learning. ACM; 2009. p. 609–16.
[86] Zou H, Hastie T. Regularization and variable selection via the elastic net. J. R. Stat. Soc. Ser. B Stat Methodol. 2005;67(2):301–20.
[87] Wold S, Esbensen K, Geladi P. Principal component analysis. Chemom. Intell. Lab. Syst. 1987;2(1–3):37–52.
[88] Shen MD, Nordahl CW, Young GS, Wootton-Gorges SL, Lee A, Liston SE, et al. Early brain enlargement and elevated extra-axial fluid in infants who develop autism spectrum disorder. Brain 2013;136(9):2825–35.
[89] Gordon EM, Laumann TO, Adeyemo B, Huckins JF, Kelley WM, Petersen SE. Generation and evaluation of a cortical area parcellation from resting-state correlations. Cereb. Cortex 2014;26(1):288–303.
[90] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. Smote: synthetic minority oversampling technique. J. Artif. Intell. Res. 2002;16:321–57.
[91] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 2014;15(1):1929–58.
[92] Mostapha M, Kim S, Wu G, Zsembik L, Pizer S, Styner M. Non-euclidean, convolutional learning on cortical brain surfaces. Biomedical imaging (ISBI 2018), 2018 IEEE 15th international symposium on. IEEE; 2018. p. 527–30.
[93] Gu X, Gortler SJ, Hoppe H. Geometry images. ACM Trans. Graph. 2002;21(3):355–61.
[94] Gulsrud AC, Hellemann GS, Freeman SF, Kasari C. Two to ten years: developmental trajectories of joint attention in children with ASD who received targeted social communication interventions. Autism Res. 2014;7(2):207–15.
[95] Kasari C, Gulsrud A, Paparella T, Hellemann G, Berry K. Randomized comparative efficacy study of parent-mediated interventions for toddlers with autism. J. Consult. Clin. Psychol. 2015;83(3):554.
[96] Drucker H, Burges CJ, Kaufman L, Smola AJ, Vapnik V. Support vector regression machines. Advances in neural information processing systems. 1997. p. 155–61.
[97] Robnik-Šikonja M, Kononenko I. Theoretical and empirical analysis of ReliefF and RReliefF. Mach. Learn. 2003;53(1–2):23–69.
[98] Mullen EM, et al. Mullen scales of early learning. MN: AGS Circle Pines; 1995.
[99] Bryson SE, Zwaigenbaum L, McDermott C, Rombough V, Brian J. The autism

observation scale for infants: scale development and reliability data. J. Autism Dev. Disord. 2008;38(4):731–8.
[100] S. S. Sparrow, D. A. Balla, D. V. Cicchetti, P. L. Harrison, E. A. Doll, Vineland adaptive behavior scales.
[101] S. Ruder, An overview of multi-task learning in deep neural networks, arXiv preprint arXiv:1706.05098.
[102] Caruna R. Multitask learning: A knowledge-based source of inductive bias. Machine learning: Proceedings of the tenth international conference. 1993. p. 41–8.
[103] Baxter J. A Bayesian/information theoretic model of learning to learn via multiple task sampling. Mach. Learn. 1997;28(1):7–39.
[104] Duong L, Cohn T, Bird S, Cook P. Low resource dependency parsing: Cross-lingual parameter sharing in a neural network parser. Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing (volume 2: Short papers). vol. 2. 2015. p. 845–50.
[105] Y. Yang, T. M. Hospedales, Trace norm regularised deep multi-task learning, arXiv preprint arXiv:1606.04038.
[106] Moeskops P, Wolterink JM, van der Velden BH, Gilhuijs KG, Leiner T, Viergever MA, et al. Deep learning for multi-task medical image segmentation in multiple modalities. International conference on medical image computing and computer-assisted intervention. Springer; 2016. p. 478–86.
[107] Suk H-I, Lee S-W, Shen D, Initiative ADN, et al. Deep sparse multi-task learning for feature selection in Alzheimer's disease diagnosis. Brain Struct. Funct. 2016;221(5):2569–87.
[108] Thung K-H, Yap P-T, Shen D. Multi-stage diagnosis of alzheimer¢s disease with incomplete multimodal data via multi-task deep learning. Deep learning in medical image analysis and multimodal learning for clinical decision support. Springer; 2017. p. 160–8.
[109] Ker J, Wang L, Rao J, Lim T. Deep learning applications in medical image analysis. IEEE Access 2018;6:9375–89.
[110] Erickson BJ, Korfiatis P, Kline TL, Akkus Z, Philbrick K, Weston AD. Deep learning in radiology: does one size fit all? J. Am. Coll. Radiol. 2018;15(3):521–6.
[111] J. Cho, K. Lee, E. Shin, G. Choy, S. Do, How much data is needed to train a medical image deep learning system to achieve necessary high accuracy?, arXiv preprint arXiv:1511.06348.
[112] Socher R, Huval B, Bath B, Manning CD, Ng AY. Convolutional-recursive deep learning for 3d object classification. Advances in neural information processing systems. 2012. p. 656–64.
[113] Snoek J, Larochelle H, Adams RP. Practical bayesian optimization of machine learning algorithms. Advances in neural information processing systems. 2012. p. 2951–9.
[114] Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, et al. Convolutional neural networks for medical image analysis: full training or fine tuning? IEEE Trans. Med. Imaging 2016;35(5):1299–312.
[115] Dolz J, Desrosiers C, Ayed IB. 3d fully convolutional networks for subcortical segmentation in MRI: a large-scale study. NeuroImage 2018;170:456–70.
[116] Xu Y, Géraud T, Bloch I. From neonatal to adult brain mr image segmentation in a few seconds using 3d-like fully convolutional network and transfer learning. 2017 IEEE international conference on image processing (ICIP). IEEE; 2017. p. 4417–21.
[117] L. Perez, J. Wang, The effectiveness of data augmentation in image classification using deep learning, arXiv preprint arXiv:1712.04621.
[118] Pereira S, Pinto A, Alves V, Silva CA. Brain tumor segmentation using convolutional neural networks in MRI images. IEEE Trans. Med. Imaging 2016;35(5):1240–51.
[119] Z. Akkus, I. Ali, J. Sedlar, T. L. Kline, J. P. Agrawal, I. F. Parney, C. Giannini, B. J. Erickson, Predicting 1p19q chromosomal deletion of low-grade gliomas from mr images using deep learning, arXiv preprint arXiv:1611.06939.
[120] Mazurowski MA, Habas PA, Zurada JM, Lo JY, Baker JA, Tourassi GD. Training neural network classifiers for medical decision making: the effects of imbalanced datasets on classification performance. Neural Netw. 2008;21(2–3):427–36.
[121] Maloof MA. Learning when data sets are imbalanced and when costs are unequal and unknown. ICML-2003 workshop on learning from imbalanced data sets II. vol. 2. 2003. (pp. 2–1).
[122] Krawczyk B. Learning from imbalanced data: open challenges and future directions. Prog. Artif. Intell. 2016;5(4):221–32.
[123] A. Janowczyk, A. Madabhushi, Deep learning for digital pathology image analysis: a comprehensive tutorial with selected use cases, J. Pathol. Inf. 7.
[124] C. J. Brown, G. Hamarneh, Machine learning on human connectome data from MRI, arXiv preprint arXiv:1611.08699.
[125] Kawahara J, Brown CJ, Miller SP, Booth BG, Chau V, Grunau RE, et al. Brainnetcnn: convolutional neural networks for brain networks; towards predicting neurodevelopment. NeuroImage 2017;146:1038–49.
[126] Brown CJ, Miller SP, Booth BG, Zwicker JG, Grunau RE, Synnes AR, et al. Predictive subnetwork extraction with structural priors for infant connectomes. International conference on medical image computing and computer-assisted intervention. Springer; 2016. p. 175–83.
[127] Brown CJ, Miller SP, Booth BG, Poskitt KJ, Chau V, Synnes AR, et al. Prediction of motor function in very preterm infants using connectome features and local synthetic instances. International conference on medical image computing and computer-assisted intervention. Springer; 2015. p. 69–76.
[128] Han H, Wang W-Y, Mao B-H. Borderline-smote: A new over-sampling method in imbalanced data sets learning. International conference on intelligent computing. Springer; 2005. p. 878–87.
[129] Jo T, Japkowicz N. Class imbalances versus small disjuncts. ACM Sigkdd Explor. Newslett. 2004;6(1):40–9.
[130] Guo H, Viktor HL. Learning from imbalanced data sets with boosting and data generation: the databoost-im approach. ACM Sigkdd Explor. Newslett. 2004;6(1):30–9.
[131] Shen L, Lin Z, Huang Q. Relay backpropagation for effective learning of deep convolutional neural networks. European conference on computer vision. Springer; 2016. p. 467–82.
[132] Haixiang G, Yijing L, Shang J, Mingyun G, Yuanyue H, Bing G. Learning from class-imbalanced data: review of methods and applications. Expert Syst. Appl. 2017;73:220–39.
[133] Kubat M, Matwin S, et al. Addressing the curse of imbalanced training sets: One-sided selection. vol. 97. Nashville, USA: Icml; 1997. p. 179–86.
[134] Barandela R, Rangel E, Sánchez JS, Ferri FJ. Restricted decontamination for the imbalanced training sample problem. Iberoamerican congress on pattern recognition. Springer; 2003. p. 424–31.
[135] Huang C, Li Y, Change Loy C, Tang X. Learning deep representation for imbalanced classification. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 5375–84.
[136] Kukar M, Kononenko I, et al. Cost-sensitive learning with neural networks. ECAI; 1998. p. 445–9.
[137] Lee H-j, Cho S. The novelty detection approach for different degrees of class imbalance. International conference on neural information processing. Springer; 2006. p. 21–30.
[138] Wang S, Minku LL, Yao X. Resampling-based ensemble methods for online class imbalance learning. IEEE Trans. Knowl. Data Eng. 2015;27(5):1356–68.
[139] Chawla NV, Lazarevic A, Hall LO, Bowyer KW. Smoteboost: Improving prediction of the minority class in boosting. European conference on principles of data mining and knowledge discovery. Springer; 2003. p. 107–19.
[140] Havaei M, Davy A, Warde-Farley D, Biard A, Courville A, Bengio Y, et al. Brain tumor segmentation with deep neural networks. Med. Image Anal. 2017;35:18–31.
[141] Ecker C, Marquand A, Mourão-Miranda J, Johnston P, Daly EM, Brammer MJ, et al. Describing the brain in autism in five dimensions magnetic resonance imaging-assisted diagnosis of autism spectrum disorder using a multiparameter classification approach. J. Neurosci. 2010;30(32):10612–23.
[142] Wang Z, Childress AR, Wang J, Detre JA. Support vector machine learning-based fmri data group analysis. NeuroImage 2007;36(4):1139–51.
[143] Haufe S, Meinecke F, Görgen K, Dähne S, Haynes J-D, Blankertz B, et al. On the interpretation of weight vectors of linear models in multivariate neuroimaging. Neuroimage 2014;87:96–110.
[144] Montavon G, Samek W, Müller K-R. Methods for interpreting and understanding deep neural networks. Digital Signal Process. 2018;73:1–15.
[145] Z. C. Lipton, The mythos of model interpretability, arXiv preprint arXiv:1606.03490.
[146] Erhan D, Bengio Y, Courville A, Vincent P. Visualizing higher-layer features of a deep network. vol. 1341 (3). University of Montreal; 2009. p. 1.
[147] K. Simonyan, A. Vedaldi, A. Zisserman, Deep inside convolutional networks: Visualising image classification models and saliency maps, arXiv preprint arXiv:1312.6034.
[148] J. Yosinski, J. Clune, A. Nguyen, T. Fuchs, H. Lipson, Understanding neural networks through deep visualization, arXiv preprint arXiv:1506.06579.
[149] A. Shrikumar, P. Greenside, A. Kundaje, Learning important features through propagating activation differences, arXiv preprint arXiv:1704.02685.
[150] L. M. Zintgraf, T. S. Cohen, T. Adel, M. Welling, Visualizing deep neural network decisions: Prediction difference analysis, arXiv preprint arXiv:1702.04595.
[151] Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. European conference on computer vision. Springer; 2014. p. 818–33.
[152] Bach S, Binder A, Montavon G, Klauschen F, Müller K-R, Samek W. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PLoS One 2015;10(7):e0130140.
[153] Ribeiro MT, Singh S, Guestrin C. Why should i trust you?: Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. ACM; 2016. p. 1135–44.
[154] Kumar D, Wong A, Taylor GW. Explaining the unexplained: A class-enhanced attentive response (clear) approach to understanding deep neural networks. IEEE computer vision and pattern recognition (CVPR) workshop. 2017.
[155] Zeiler MD, Taylor GW, Fergus R. Adaptive deconvolutional networks for mid and high level feature learning. Computer vision (ICCV), 2011 IEEE international conference on, IEEE. 2011. p. 2018–25.
[156] F. Doshi-Velez, B. Kim, Towards a rigorous science of interpretable machine learning, arXiv preprint arXiv:1702.08608.
[157] Caruana R, Lou Y, Gehrke J, Koch P, Sturm M, Elhadad N. Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining. ACM; 2015. p. 1721–30.
[158] Elsayed G, Shankar S, Cheung B, Papernot N, Kurakin A, Goodfellow I, et al. Adversarial examples that fool both computer vision and time-limited humans. Advances in neural information processing systems. 2018. p. 3914–24.
[159] Nguyen A, Yosinski J, Clune J. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. p. 427–36.
[160] Bernardo J, Bayarri M, Berger J, Dawid A, Heckerman D, Smith A, et al. Generative or discriminative? Getting the best of both worlds. Bayesian Stat. 2007;8(3):3–24.
[161] Bengio Y, Laufer E, Alain G, Yosinski J. Deep generative stochastic networks trainable by backprop. International conference on machine learning. 2014. p. 226–34.
[162] Hinton GE, Osindero S, Teh Y-W. A fast learning algorithm for deep belief nets.

Neural Comput. 2006;18(7):1527–54.
[163] Srivastava N, Salakhutdinov RR. Multimodal learning with deep Boltzmann machines. Advances in neural information processing systems. 2012. p. 2222–30.
[164] Hinton GE. Deep belief networks. Scholarpedia 2009;4(5):5947.
[165] D. P. Kingma, M. Welling, Auto-encoding variational bayes, arXiv preprint arXiv:1312.6114.
[166] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. Advances in neural information processing systems. 2014. p. 2672–80.
[167] Dosovitskiy A, Brox T. Generating images with perceptual similarity metrics based on deep networks. Advances in neural information processing systems. 2016. p. 658–66.
[168] A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, arXiv preprint arXiv:1511.06434.
[169] M. Mirza, S. Osindero, Conditional generative adversarial nets, arXiv preprint arXiv:1411.1784.
[170] Li C, Wand M. Precomputed real-time texture synthesis with markovian generative adversarial networks. European conference on computer vision. Springer; 2016. p. 702–16.
[171] J.-Y. Zhu, T. Park, P. Isola, A. A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, [arXiv preprint].
[172] A. Odena, C. Olah, J. Shlens, Conditional image synthesis with auxiliary classifier gans, arXiv preprint arXiv:1610.09585.
[173] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein gan, arXiv preprint arXiv:1701.07875.
[174] Mao X, Li Q, Xie H, Lau RY, Wang Z, Smolley SP. Least squares generative adversarial networks. Computer vision (ICCV), 2017 IEEE international conference on. IEEE; 2017. p. 2813–21.
[175] S. Kazeminia, C. Baur, A. Kuijper, B. van Ginneken, N. Navab, S. Albarqouni, A. Mukhopadhyay, Gans for medical image analysis, arXiv preprint arXiv:1809.06222.
[176] X. Yi, E. Walia, P. Babyn, Generative adversarial network in medical imaging: A review, arXiv preprint arXiv:1809.07294.
[177] Frid-Adar M, Klang E, Amitai M, Goldberger J, Greenspan H. Synthetic data augmentation using gan for improved liver lesion classification. Biomedical imaging (ISBI 2018), 2018 IEEE 15th international symposium on. IEEE; 2018. p. 289–93.
[178] Chuquicusma MJ, Hussein S, Burt J, Bagci U. How to fool radiologists with generative adversarial networks? A visual turing test for lung cancer diagnosis. Biomedical imaging (ISBI 2018), 2018 IEEE 15th international symposium on. IEEE; 2018. p. 240–4.
[179] Bermudez C, Plassard AJ, Davis LT, Newton AT, Resnick SM, Landman BA. Learning implicit brain mri manifolds with deep learning. Medical imaging 2018: Image processing. vol. 10574. International Society for Optics and Photonics; 2018. p. 105741L.
[180] Wolterink JM, Dinkla AM, Savenije MH, Seevinck PR, van den Berg CA, Išgum I. Deep mr to ct synthesis using unpaired data. International workshop on simulation and synthesis in medical imaging. Springer; 2017. p. 14–23.
[181] Q. Yang, N. Li, Z. Zhao, X. Fan, E.-C. Chang, Y. Xu, et al., MRI image-to-image translation for cross-modality image registration and segmentation, arXiv preprint arXiv:1801.06940.
[182] Vemulapalli R, Van Nguyen H, Kevin Zhou S. Unsupervised cross-modal synthesis of subject-specific scans. Proceedings of the IEEE international conference on computer vision. 2015. p. 630–8.
[183] Huang Y, Shao L, Frangi AF. Simultaneous super-resolution and cross-modality synthesis of 3d medical images using weakly-supervised joint convolutional sparse coding. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 6070–9.
[184] Xue Y, Xu T, Zhang H, Long LR, Huang X. Segan: adversarial network with multi-scale l 1 loss for medical image segmentation. Neuroinformatics 2018:1–10.
[185] Moeskops P, Veta M, Lafarge MW, Eppenhof KA, Pluim JP. Adversarial training and dilated convolutions for brain mri segmentation. Deep learning in medical image analysis and multimodal learning for clinical decision support. Springer; 2017. p. 56–64.
[186] Zhao M, Wang L, Chen J, Nie D, Cong Y, Ahmad S, et al. Craniomaxillofacial bony structures segmentation from mri with deep-supervision adversarial learning. International conference on medical image computing and computer-assisted intervention. Springer; 2018. p. 720–7.
[187] X. Chen, E. Konukoglu, Unsupervised detection of lesions in brain mri using constrained adversarial auto-encoders, arXiv preprint arXiv:1806.04972.
[188] C. Baur, B. Wiestler, S. Albarqouni, N. Navab, Deep autoencoding models for unsupervised anomaly segmentation in brain mr images, arXiv preprint arXiv:1804.04488.
[189] Baumgartner CF, Koch LM, Tezcan KC, Ang JX, Konukoglu E. Visual feature attribution using Wasserstein gans. Proc IEEE comput soc conf comput vis pattern recognit. 2017.
[190] Schlegl T, Seeböck P, Waldstein SM, Schmidt-Erfurth U, Langs G. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. International conference on information processing in medical imaging. Springer; 2017. p. 146–57.
[191] Zhang L, Gooya A, Frangi AF. Semi-supervised assessment of incomplete lv coverage in cardiac mri using generative adversarial nets. International workshop on simulation and synthesis in medical imaging. Springer; 2017. p. 61–8.