Classification of Autism Spectrum Disorder from Structural and Functional Magnetic Resonance Imaging Using Machine Learning

Andrew Bender, Vanderbilt University

Mentors: Dr. Mark Wallace & David Tovar

Thesis Committee: Dr. Mark Wallace, Dr. Ramnarayan Ramachandran,

Dr. Randolph Blake, and Dr. Eli Zimmerman

April 12, 2018

Table of Contents

Chapter 1: Introduction to Classification and Machine Learning

Classification

Classification is a class of problems within machine learning that is concerned with identifying which of a set of categories a new observation belongs to. Classifiers learn to predict which class an observation belongs to through training with a set of observations for which the class of each observation is known (Kotsiantis, Zaharakis, & Pintelas, 2006). This learning is considered "supervised" because the classifier knows what class each observation belongs to and uses this knowledge to perform classification. Clustering is the corresponding "unsupervised" machine learning procedure; in clustering, unlabeled data is grouped into categories based on some measure of inherent similarity or distance (Grira, Crucianu, & Boujemaa, 2004). The set of observations which the classifier is trained on is referred to as the training set. Each observation has a set of features that define it. The performance of a classifier is evaluated using a test set, which contains observations that the classifier was not used to train on. The classifier's classification accuracy is simply the number of observations the classifier classifies correctly in the test set divided by the total number of observations in the test set. Figure 1 contains 200 observations from Group X and 200 observations from Group Y. Each of these observations has two features. These two features, which

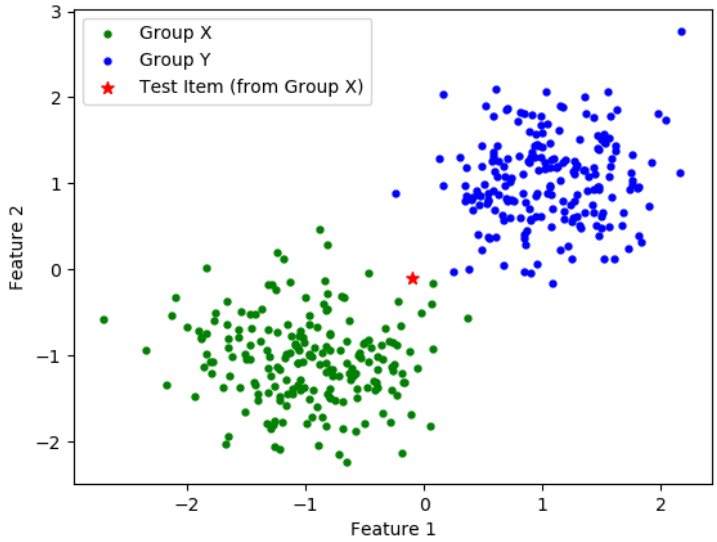

Figure 1. An example classification problem. The classifier would be trained with data from Groups X and Y to identify whether a test item is from Group X or Group Y based on the two features of the test item.

each have a numerical value, comprise the feature vector for the observation and define a point in two-dimensional space that represents the observation. This two-dimensional space is called the feature space. These ideas extend to circumstances in which each observation has many more features; when the observations have *n* features, each observation will have a *n*-dimensional feature vector that defines a point in the *n*-dimensional feature space that represents the observation. The remainder of this chapter is dedicated to various classification methods, each one having its own advantages and disadvantages and level of sophistication. This overview of classification methods is not meant to be comprehensive; rather, it serves as a basic introduction to the topic of classification and to the particular classifiers used in this paper for classification of autism spectrum disorder.

K-Nearest Neighbors

The k-nearest neighbors (kNN) classifier is one of the simplest classification algorithms. There is no training required, as kNN is a type of instance-based learning (also called lazy learning), where all computation is deferred until classification of the test set. For each item in the test set, its Euclidean distance from every item in the training set is calculated. When k = 1, the algorithm is referred to as the nearest neighbor algorithm, as each test item is assigned the label of its nearest neighbor in the multidimensional feature space (i.e.

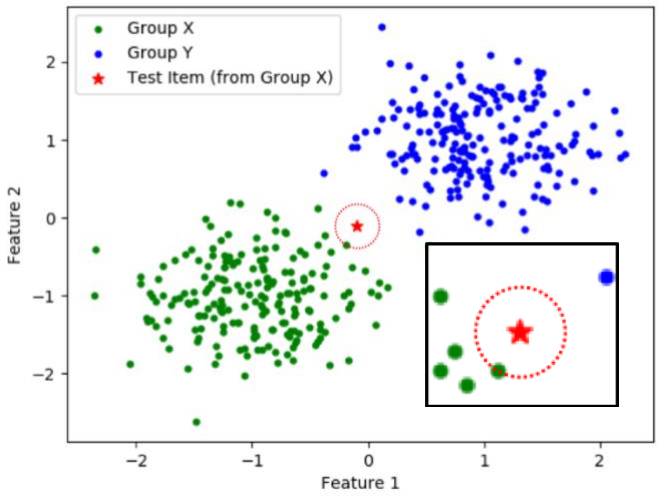

Figure 2. Illustration of how the k-nearest neighbors algorithm works when k = 1. A zoomed-in view of the test item is shown in the bottom right.

the item that has the smallest Euclidean distance) (Zhang, 2016). In Figure 2, for example, the item with the smallest Euclidean distance in the two-dimensional feature space from the test item is an item from Group X, so the test item will be predicted to be from Group X. When k > 1, there are two primary ways in which the class label can be determined: (1) majority vote and (2) weighting. In majority voting, the class with the most items of the k-nearest neighbors to the test item is the class label given to the test item. In weighting, each of the k-nearest neighbors to the test item is given a weight of 1/(Euclidean distance from the test item) and items from the same class are summed up. The class with the highest sum of weights is then the class label given to the test item (Coomans & Massart, 1982). Although kNN is a simple algorithm that makes no assumptions about the data, it can be computationally expensive when the data is complex or when k is large. kNN is also sensitive to irrelevant features, especially when k is small, as there is no training to determine which features are relevant for classification; the k-nearest neighbors are used to determine the classifier prediction regardless of how relevant those neighbors really are.

Naïve-Bayes

The Naïve-Bayes classifier is a classification algorithm based on Bayes' Theorem. Bayes' Theorem gives the probability that event A will occur given that event B already occurred (denoted as P(A|B)) through the following formula:

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

In the case of classification, A is the class labels and B is the features of the data. P(A|B), also referred to as the posterior probability, is the conditional probability of a particular class label given a certain set of features. Likewise, P(B|A) is the conditional probability of a certain set of

features given a particular class label. Finally, P(A) is the probability of the particular class label and P(B) is the total probability of the features. Because the value of each of the features is given and does not depend on the class labels, P(B) is effectively constant and ignored by the classification algorithm. For the Naïve-Bayes classifier, we assume that the features are conditionally independent, such that:

$$P(B|A) = P(B_1|A) * P(B_2|A) * ... * P(B_n|A)$$

The assumption that the features are independent rarely is true in real-world data (hence the use of the term naïve), but it enables straightforward computation and often produces superior results to more sophisticated methods that do not assume independence (Hand & Yu, 2001; Rish, 2001). Researchers have posited that this unexpected discrepancy is due to the low-variance probability estimates that result from the intrinsic simplicity of the naïve approaches (Hand & Yu, 2001; Rish, 2001). The probabilities of the features given each class label (i.e. $P(B_1|A)$, $P(B_2|A)$, … , $P(B_n|A)$) are assumed to be from a Gaussian probability distribution when the features are continuous (John & Langley, 1995). Each feature's mean and variance are computed across the training set for each class and then are used to define a Gaussian probability distribution for that feature given each class label. To predict the class of one test item, the Naïve-Bayes classifier does the following for each class label:

1. Uses the value of each feature as the input to that feature's Gaussian probability distribution for that particular class label.
2. Takes the outputs of each feature's probability distribution to be the conditional probability of that feature given the class label.
3. Multiplies all the feature's conditional probabilities by each other and by the probability of the class label, which is simply the fraction of the training set that is from that

particular class. The result is the posterior probability of that class label for that set of features (i.e. for the test item).

When the posterior probability for all class labels given the features of the test item have been calculated, the class label with the highest posterior probability, which is known as the *maximum a posteriori*, is considered the predicted class of the Naïve-Bayes classifier (Herman, De Pierro, & Gai, 1992). Because Naïve-Bayes assume conditional independence, it is unable to incorporate feature interactions within the probabilistic model it uses for classification, limiting the information it can draw from the data to make an accurate class prediction. Despite the naïve assumption Naïve-Bayes classifiers make, the classifier scales well with large numbers of features, is simple and easy to implement, works well with small amounts of training data, and is not sensitive to irrelevant features like kNN is.

Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) aims to find a linear combination of features that separates two or more classes. In order to optimize the separation between the classes, LDA maximizes the ratio of the distance between the class means to the variance of the classes (Maroco et al., 2011). LDA makes some assumptions about the data in order to simplify the mathematics and make the algorithm computationally feasible (Büyüköztürk & Çokluk-Bökeoğlu, 2008):

1. Each feature is normally distributed across the items for each class.
2. The variance of items across classes is the same.
3. Items within a class are assumed to be independent.

Although many datasets violate one or more of these assumptions, it has been shown that LDA is relatively robust to slight violations of these assumptions (Lachenbruch & Goldstein, 1979).

When each item in the dataset has only two features, LDA provides a line that best separates the two features. This line can then be used as the decision boundary for classification. In the example shown to the right, any test item determined to be above the decision boundary in the feature space would be classified as Group Y, while any test item below the decision boundary would be classified as Group X. Thus, the test item shown in Figure 3 would be classified as Group Y. These principles also apply to higher dimensional feature spaces and classification problems involving more than two classes.

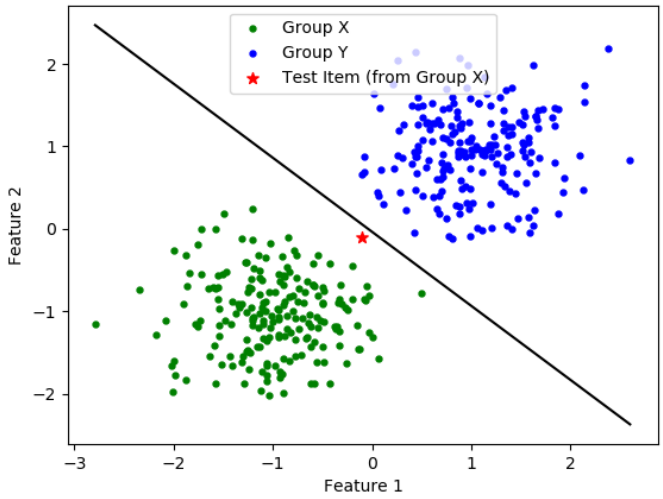

Figure 3. Linear decision boundary determined by LDA from a linearly separable training set.

In the example shown in Figure 3, the two datasets are linearly separable, meaning there is at least one line in the two-dimensional feature space that fully separates the two classes, and all the assumptions made by LDA are true of the data. Figure 4 shows one case in which the data is not linearly separable and violates Assumption #1 from above. Because the data is not

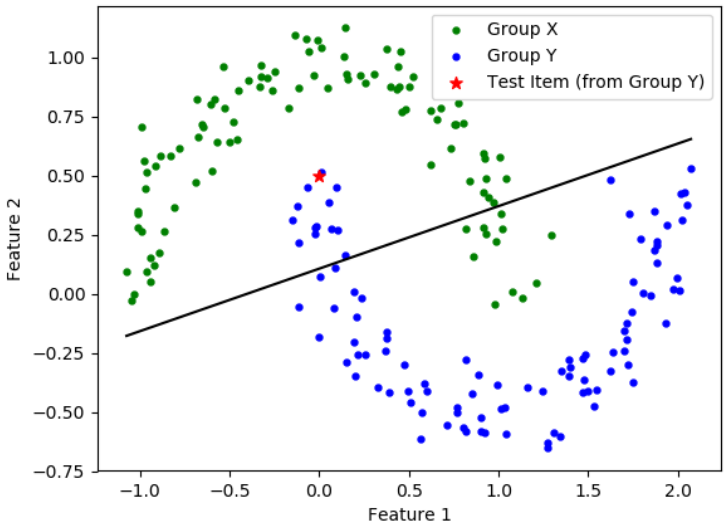

Figure 4. Linear decision boundary determined by LDA from a training set that is not linearly separable.

linearly separable, some items from Group X are below the decision boundary and would be mistakenly predicted to be from Group Y by the LDA classifier, while some items from Group Y are above the decision boundary and would be mistakenly predicted to be from Group X. The test item displayed in Figure 4 is one example of an incorrect prediction made by the classifier; although it is from Group Y, the classifier will predict that it is from Group X because it is above the linear decision boundary. The effectiveness of LDA, then, is dependent on the data it uses for classification; LDA performs well with data that is linearly separable and can struggle with classification of data that is only nonlinearly separable.

Artificial Neural Networks

Artificial neural networks are a branch of machine learning involving a network of interconnected units called nodes that are loosely modeled after neurons in the brain (Hopfield, 1988; Rumelhart, Hinton, & McClelland, 1986). Neural networks learn to perform tasks through presentation of labeled examples and application of learning rules that change the network's structure (Rumelhart et al., 1986). During training, labeled examples serve as input to the neural network and the signal is propagated through the neural network. The error between the output of the neural network and the label is used to calculate modifications to the neural network's weights through a process known as backpropagation. These modifications are called gradients and backpropagation uses the chain rule to propagate these gradients from the output of the neural network backwards through all the neural network's layers. Through repeated training with labeled examples, the weights of the neural network become tuned so that the neural network can map each input (i.e. labeled example) to the correct output (i.e. label or classification).

Fully Connected Neural Networks

The prototypical neural network is the fully connected neural network, in which each node in one layer is connected to every node in both the preceding and following layers. Each of these connections has a particular weight, which is synonymous with the synaptic strength between two neurons. Each node also has an activation function, which dictates how the net input from all the nodes in the preceding layer is mapped to the node's activation. The simplest activation function is a linear activation function, where the activation of the node is set to be equal to the node's net input. The rectified linear unit (ReLU) activation function, which performs the same as the linear activation function for positive net inputs and maps all negative net inputs to an activation of zero, has become increasingly popular in recent years (Nair & Hinton, 2010). Because of the sparse activation it causes within the neural network, use of ReLU activation functions ease training and improve network classification performance. Batch normalization operations are also commonly used in neural networks to normalize the input or output of a layer to have zero mean and unit variance, which helps prevent the network from diverging during training (Ioffe & Szegedy, 2015). When used for image classification, fully connected neural networks have a few major drawbacks (LeCun, Haffner, Bottou, & Bengio, 1999). First, fully connected networks have no built-in invariance to translation, scaling, or rotation of the image, requiring images to be size-normalized and centered for image classification to be successful. Second, because the layers of the neural networks are fully connected, the number of trainable parameters (i.e. weights) within the network grows quickly as the number of pixels in the image increases. A large number of trainable parameters makes training computationally expensive and renders the network susceptible to overfitting, which is when the network overlearns the images it is trained on at the expense of generalizability.

Finally, the topology of the image is entirely ignored by fully connected neural networks. The image has to be flattened to one dimension, so its two-dimensional structure is ignored, which further confounds image classification.

Convolutional Neural Networks

Convolutional neural networks (CNNs) were designed to address the difficulties faced by fully connected neural networks in image classification. Nodes within a CNN perform convolution and pooling operations that mimic the behavior of simple and complex cells in primary visual cortex (Fukushima, 1980; LeCun et al., 1999). Convolution operations process data within a local receptive field, functioning like V1 simple cells; pooling operations then combine the output from convolution operations together like a V1 complex cell does with the output from a collection of V1 simple cells. Each node serves as a feature detector within the local region that is its receptive field. The output of each node is referred to as a feature map and serves as the input for following convolutional layers. The weights of one node are shared with other nodes within a convolutional layer, such that the same feature detection is done at each portion of the image, enabling some degree of translation invariance. This weight sharing also reduces the number of trainable parameters substantially, making training more computationally efficient and mitigating the network's susceptibility to overfitting. The convolution and pooling operations can be performed multi-dimensionally, enabling CNNs to account for the topology of images.

Putting convolutional layers one after another allows for increasingly complex features to be extracted from the image, as elementary features from early layers are combined by subsequent layers to detect higher-order features. Deep CNNs, in which many convolutional

layers are placed in series, one after another, have become state-of-the-art in image classification tasks (Krizhevsky, Sutskever, & Hinton, 2012; Simonyan & Zisserman, 2014; Szegedy et al., 2014), in some cases even surpassing human level performance (He, Zhang, Ren, & Sun, 2015b). Various efforts have been made to visualize the activations of CNNs in order to understand how they perform image classification (Fong & Vedaldi, 2017; Ribeiro, Singh, & Guestrin, 2016; Selvaraju et al., 2017; Yang, Rangarajan, & Ranka, 2018a; Zhou, Khosla, Lapedriza, Oliva, & Torralba, 2016). These efforts have shown that CNNs extract rudimentary features like edges in earlier layers regardless of the image classification task, and extract complex features relevant to the image classification task in later layers (i.e. whiskers for a task involving classification of a cat). Because backpropagation uses the chain rule, gradients can become very small in the beginning layers of the neural network when the neural network is deep. As a result, deep CNNs are susceptible to the vanishing gradient problem, where the gradients calculated for weight modification become so small that the neural network can no longer be trained properly (He, Zhang, Ren, & Sun, 2015a). With gradients close to zero, the neural network no longer changes during training and cannot learn any longer, rendering it useless.

The fundamental building block of a CNN is the convolutional layer. Convolutional layers transform their input by utilizing the convolution operation (see Figure 5 on the following page). A convolutional layer has a distinct number of filters, which serve as feature detectors, each filter having a particular feature it detects. Each filter is presented at local regions throughout the input image. At each local region, the "alignment" of the filter with the image at that local region is calculated. Effectively, the convolution of the image with a filter calculates the alignment of the filter with every local region in the image. The net result of the convolution operation, then, is an array of how well-aligned the filter and image are at every local region

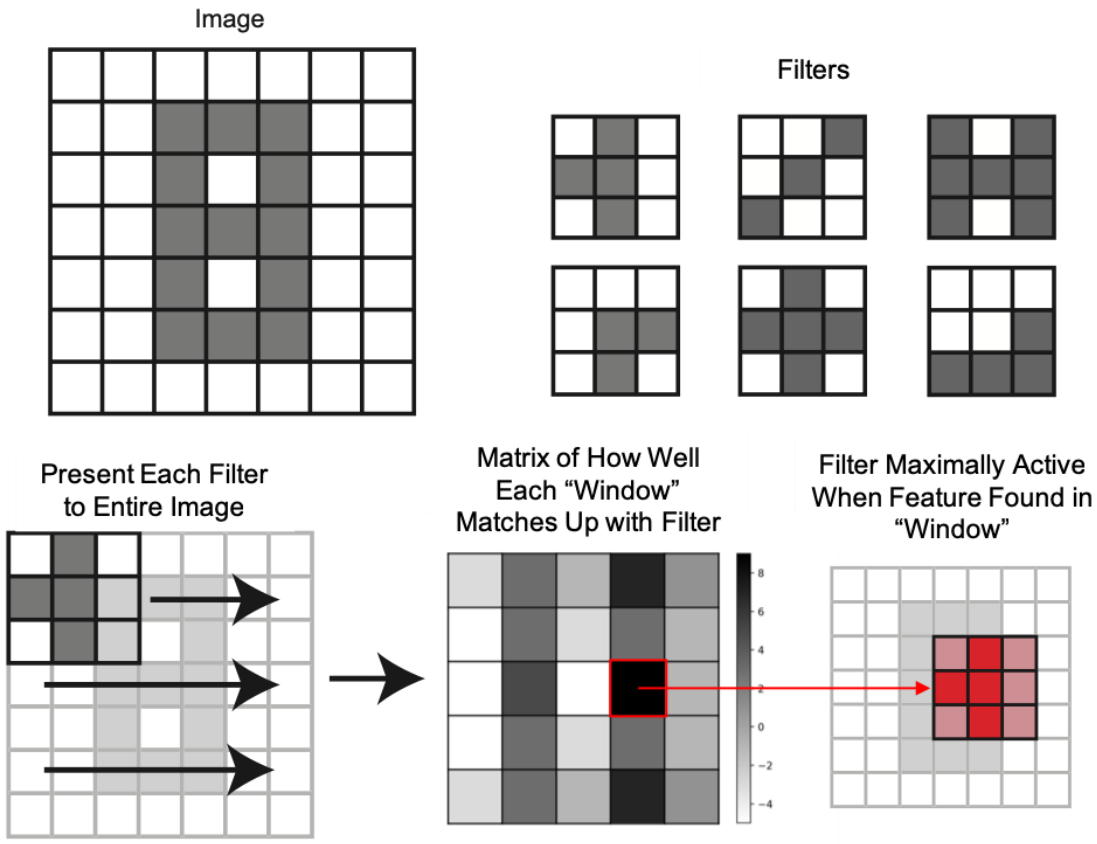

Figure 5. Convolutional Layers and Filters

within the image, as shown in Figure 5. This array is called a feature map. In a convolutional layer, the image is convolved with each filter of the convolutional layer. The output of a convolution layer is a collection of feature maps, one for convolution of the input image with each filter.

In CNNs, convolutional layers are often either preceded or immediately followed by pooling operations. Pooling operations reduce the spatial size of the network, simplifying the network's activations. There are two primary types of pooling operations used in neural network architectures: max pooling and average pooling (see Figure 6 on the following page). In max pooling, the maximum value of each pool is the value taken by the unit representing the pool in the output of the pooling operation. Max pooling is generally used at the beginning of a CNN because it extracts the most highly active areas of the input image, eliminating unnecessary,

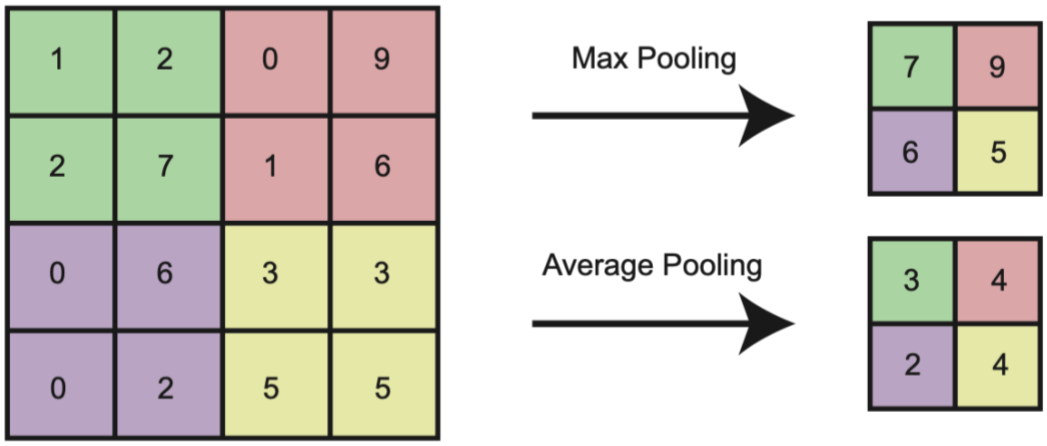

Figure 6. Pooling Operations

insignificant data from the image and reducing the input down to what is essential. By contrast, in average pooling, the average of the values within a pool serves as the activation for the unit representing the pool in the output of the pooling operation. Average pooling is often done at the end of a CNN because it helps to reduce the spatial size of the networks while taking all the values within the pool into account. The benefits of pooling are two-fold. First, the smaller spatial size lessens the number of computations required by subsequent layers. Second, the reduction in the number of parameters within the neural network makes the network less prone to overfitting.

Residual Neural Networks

Inspired by skip connections within the brain, residual neural networks are a variants of CNNs that have shortcut connections from the beginning of a convolutional layer to the end of a convolutional layer (He et al., 2015a; He, Zhang, Ren, & Sun, 2016). These shortcut connections can involve skipping of multiple convolutional layers, in which case all the layers in between the beginning and end of the shortcut connection are called a residual block (see Figure 7 on the following page). In the simplest case, the shortcut connections are an identity mapping, such that the input to a convolutional layer is just added to the output of that convolutional layer.

Because the identity mapping cannot be changed during training, the only portion of the neural network being modified is the convolutional layer. Training, then, is only optimizing a residual portion of the network, hence the name residual neural networks. Importantly, the shortcut connections alleviate many of the issues faced by deep CNNs, such as the vanishing gradient problem (He et al., 2015a), by allowing important features from early layers to be passed directly to later layers without having to be maintained through convolutional layers, reducing the effective depth of the network while still enabling further feature extraction with convolutional layers. Visualization techniques that are used for CNNs can also be used with residual neural networks (Yang, Rangarajan, & Ranka, 2018b). The residual neural networks constructed in this paper utilize convolutional layers and pooling operations together with residual skip connections to extract high-level features from the input MRI images and classify a complex disorder.

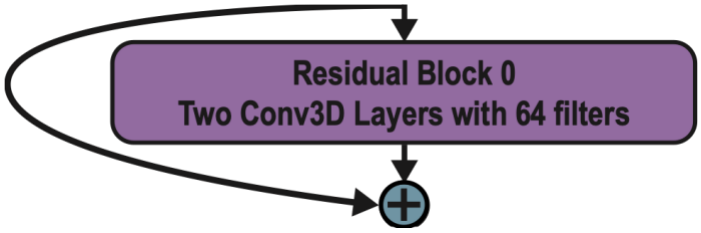

Figure 7. Example of a Residual Block

Neural Networks in Medical Imaging

Convolutional neural networks (CNNs) have been used in medical imaging applications since the 1990s in areas such as lung structure and nodule detection and breast tissue classification, although computational limitations and a lack of labeled training data stymied deep CNNs from being utilized (Bernal et al., 2017). Deep CNNs are now used for a host of medical imaging problems (see Table 1 on the following page) and are utilized widely in brain MRI for lesion detection and segmentation, as well as tumor segmentation (Bernal et al., 2017; Pereira, Pinto, Alves, & Silva, 2016). 3D CNN architectures have the potential to fully leverage

the topology of 3D volumes like structural MRI. However, because of both ease and computational considerations, most of the medical imaging applications of CNNs use two-dimensional slices of the brain as input, utilizing the well-developed 2D CNN architectures used for natural image classification. Those medical imaging applications that do utilize 3D CNNs usually extract portions of the MRI for use as input to the CNN. While 3D CNN architectures are more computationally expensive than 2D architectures, recent advances in computational power have enabled usage of whole brain MRIs as input to a 3D CNN that predicts Alzheimer's disease (Payan & Montana, 2015; Yang et al., 2018b). In fact, 3D convolutional architectures have outperformed 2D convolution architectures in patch-based segmentation using the ADNI hippocampus MRI dataset (Lai, 2015). Fortunately, the visualization techniques used for 2D CNNs have also been extended to 3D CNNs and have been used to show that the cerebral cortex and lateral ventricle, two areas believed to be important for Alzheimer's disease diagnosis by physicians, were particularly important for the CNN's correct classification of Alzheimer's disease (Yang et al., 2018b).

Table 1. Medical Imaging Applications of Neural Networks

Disease	Accuracy	Architecture	Study
Interstitial lung diseases	85.5%	2D CNN	(Anthimopoulos, Christodoulidis, Ebner, Christe, & Mougiakakou, 2016)
Alzheimer's disease	79.4%	3D CNN	(Yang et al., 2018b)
Myocardial infarction	93.53%	2D CNN	(Acharya, Fujita, Oh, et al., 2017)
Seizures	88.67%	2D CNN	(Acharya, Oh, Hagiwara, Tan, & Adeli, 2018)
Coronary Artery Disease	94.95%	2D CNN	(Acharya, Fujita, Lih, et al., 2017)
Parkinson's disease	88.25%	2D CNN	(Oh et al., 2018)
Brain tumor	86.5%	2D CNN	(Havaei et al., 2017)
Thyroid disease	89.29%	Fully connected neural network with multiple layers	(Ozyilmaz & Yildirim, 2002)
Heart disease	89.01%	Ensemble of multi-layered fully connected neural networks	(Das, Turkoglu, & Sengur, 2009)

Chapter 2: Classification of Autism Spectrum Disorder Using Machine Learning

Abstract

Machine learning and deep neural networks have begun to show promise in helping to diagnose diseases such as Alzheimer's disease from neuroimaging datasets. Given the utility machine learning has already shown in diagnosing diseases, here we aim to develop machine learning classifiers, including a deep residual neural network called ASDNet, that can classify autism spectrum disorder (ASD) from neuroimaging data. Because early diagnosis and early treatment can improve quality of life for individuals with ASD and their families and current diagnostic measures can only be used with children 12 months and older, machine learning shows promise for enabling earlier diagnosis and treatment of ASD. We trained linear discriminant analysis (LDA), Naïve-Bayes, and nearest neighbor classifiers, as well as ASDNet, on structural MRI, cortical thickness, and functional MRI volumes, achieving classification accuracies of 55%, 60%, and 59%, respectively. Ensembles from multiple imaging types improved classification accuracy above that achieved from use of a single imaging type, indicating that examination of differences between ASD and typically developed controls in multiple imaging types is important for effective classification. LDA was the most effective classifier across imaging types, although there is reason to believe that ASDNet's classification performance could be much improved with validation of its implementation and effective selection of the parameters that dictate how it is trained. Finally, distance to bound analysis showed that structural, cortical thickness, and functional differences in ASD extracted by the classifier are linked to metrics from behavioral assessments of ASD like the Autism Diagnostic Observation Schedule (ADOS).

Introduction

Autism spectrum disorder (ASD) is a developmental disorder characterized by early-emerging social and communication impairments along with rigid and repetitive patterns of behavior and interest (Frith & Happé, 2005). Manifestation of social impairments and rigid, repetitive behavior varies greatly with age and ability. Currently, pediatricians and psychiatrists diagnose autism spectrum disorders (ASD) using clinical criteria from the Diagnostic and Statistical Manual of Mental Disorders (DSM-5). The accepted gold-standard diagnostic assessments for these clinical criteria are the Autism Diagnostic Observation Schedule (ADOS) (Lord et al., 2000) and the Autism Diagnostic Interview-Revised (ADI-R) (Lord, Rutter, & Couteur, 1994). ADOS is a semi-structured, clinician-administered interview that assesses a variety of characteristics specific to ASD: verbal and non-verbal communication, reciprocal interaction, and conversational ability, while the ADI-R is a standardized, semi-structured, investigator-based interview for caregivers of autistic individuals. Both tools assess the same clinical criteria and show reliable specificity and sensitivity for ASD (Falkmer, Anderson, Falkmer, & Horlin, 2013), although ADOS does have a higher sensitivity than ADI-R (Randall et al., 2018). Joint administration of ADOS and ADI-R was no more accurate than use of ADOS alone (Randall et al., 2018). Because the ways in which one can test for the deficits found in ASD vary based on an individual's capabilities, there are separate ADOS modules based on age, language, and developmental level (Brentani et al., 2013). The Toddler Module of ADOS-2 has demonstrated 97% inter-rater reliability as well as 91% sensitivity and 91% specificity in diagnostic classification of ASD for children as young as 12 months (Esler, Amy N; Bal, Vanessa H; Guthrie, Whitney; Weismer, Susan E; Lord, 2015; Luyster et al., 2009).

Studies have shown that accurate and early diagnosis and treatment of ASD is critical in improving not just the quality of life of individuals with ASD, but also those of their families (Elder, Kreider, Brasher, & Ansell, 2017). Beyond the quality of life enhancement that early diagnosis and treatment provides, it has been estimated that early diagnosis and intervention provide a healthcare savings of $208,500 per child (Chasson, Harris, & Neely, 2007). Children who receive early intensive behavioral treatment have been shown to make substantial, sustained gains in IQ, language, academic performance, and adaptive behavior, and their outcomes have been significantly better than those of children in control groups (Myers & Johnson, 2007). Given the importance of early diagnosis and the fact that current diagnostic tools (e.g. ADOS) cannot be used with children younger than 12 months, it is critical that other diagnostic tools be developed.

Interventions can be started soon after diagnosis of ASD. Myers & Johnson (2007) consider there to be seven important principles of effective early childhood intervention for children with ASD (see Table 2). There are various methods that can be used with children with ASD to address the above principles. Applied behavior analysis (ABA) is based on findings from experimental psychology research used to increase and maintain desirable adaptive behaviors, reduce interfering maladaptive behaviors, teach new

Table 2. Principles of Effective Early Childhood Invention for ASD

1. Active engagement of the child at least 25 hours per week, 12 months per year, in planned, developmentally appropriate educational activities designed to address identified objectives
2. Low student-to-teacher ratio
3. Inclusion of parent training
4. Promotion of interaction with typically developing peers
5. Ongoing measurement and documentation allowing for adjustments to program when necessary
6. Incorporation of a high degree of structure
7. Use of assessment-based curricula that address communication and social skills, functional skills that prepare child for increased responsibility and independence, disruptive or maladaptive behavior, cognitive skills, and traditional readiness skills and academic skills

skills, and generalize behaviors to new environments or situations (Myers & Johnson, 2007). While interventions for ASD are usually focused on early childhood, there is empirical support for use of certain educational strategies, particularly those that are based on ABA, across all age groups for individuals with ASD (Myers & Johnson, 2007). ASD interventions are assessed through evaluation of the individuals progress in (1) improving his/her communication and social skills, cognitive skills, and traditional readiness and academic skills, (2) eliminating or reducing maladaptive or disruptive behavior, and (3) cultivating functional skills that enable the individual to handle increased responsibility and independence. Usually this assessment is done by pediatricians or psychiatrists as a person progresses through interventions.

At this point, ASD is diagnosed solely through behavioral testing using assessment tools like ADOS and ADI-R. Decisions about neuroimaging in children with A SD are made on a case-by-case basis and yield of MR I is low in children with ASD and no other neurologic findings (Battaglia & Carey, 2006; Cooper et al., 2016). While some efforts have been made to use neuroimaging to aid in diagnoses by examining for changes in brain network organization (Chen, Jiao, & Herskovits H., 2011), the success of these efforts is largely dependent on experts to manually delineate brain areas. Neuroimaging studies increasingly support a picture of early atypical brain development and widespread alterations in functional and structural connectivity throughout the brain in individuals with ASD that is thought to occur because of abnormal synaptic pruning and neural apoptosis (Muhle, Reed, Stratigos, & Veenstra-VanderWeele, 2018). The fact that a majority of the mutated genes converge on specific biological pathways that influence synaptic plasticity and connectivity at different levels further supports this theory of atypical development caused by abnormal synapse pruning (Bourgeron, 2015). Recent neuroimaging studies with high-risk infants have suggested that infants who will later develop

ASD already exhibit abnormal connectivity, particularly in regions involved in low-level sensory processing, at six months of age (Emerson et al., 2017; Lewis et al., 2017; Wolff, Jacob, & Elison, 2018; Wolff et al., 2012). Starting between the ages of six and 12 months, children who will eventually be diagnosed with ASD show increased cortical surface area, starting in sensory domains underlying auditory and visual processing, followed by more global overgrowth throughout the brain from the ages of 12 to 24 months (Hazlett et al., 2017; Shen et al., 2013). From the ages of 2 to 4 years, children with ASD continue to exhibit enlarged brain volumes compared with those of their peers (Courchesne, Campbell, & Solso, 2011; Redcay & Courchesne, 2005; Chen et al., 2011). Brain volumes of typically developing children converge with those in children with ASD by school age (Courchesne et al., 2011; Redcay & Courchesne, 2005), although children with ASD continue to have increased amygdala volumes until adolescence (Amaral, Schumann, & Nordahl, 2008; Chen et al., 2011).

While there is a great heterogeneity in both the functional and structural differences in ASD (Chen et al., 2011; Lenroot & Yeung, 2013), fMRI has shown that functional underconnectivity throughout the brain appears to be a universal characteristic of the brain in ASD (Williams, 2007). Different studies have found a variety of brain areas to be functionally under-connected in ASD, as shown in Table 3 on the following page. While the brain, as a whole, exhibits functional underconnectivity in ASD, there are regions are the brain that do show increased functional connectivity. For example, resting-state functional magnetic resonance imaging (fMRI) analysis performed in individuals with ASD found both a hypoconnectivity of long cortical-cortical and interhemispheric projections and a hyperconnectivity of local connections in subcortical regions compared with age-matched, typically developing controls (Di Martino et al., 2014). Similarly, children and adolescents with ASD display reduced short

Table 3. Functional Differences Between ASD and TDA

Brain Areas/Networks	Finding	Resting state or task-dependent?	Study
Ventral medial prefrontal cortex	Reduced activity in judgment and resting conditions	Both	(Kennedy & Courchesne, 2008a)
Ventral medial prefrontal cortex	Reduced activity in judgment and resting conditions	Both	(Kennedy & Courchesne, 2008a)
	Activity correlated with clinical measure of social impairment	Resting state	(Kennedy, Redcay, & Courchesne, 2006)
Resting-state network	Functional underconnectivity	Resting state	(Cherkassky, Kana, Keller, & Just, 2006)
	No deactivation during cognitively demanding tasks	Task-dependent	(Kennedy et al., 2006)
	Altered functional organization	Resting state	(Kennedy & Courchesne, 2008b)
Anterior-posterior connections	Functional underconnectivity	Resting state	(Cherkassky et al., 2006)
Fronto-parietal connections	Connectivity was correlated with size of corpus callosum in ASD group only	Task-dependent	(Just, Cherkassky, Keller, Kana, & Minshew, 2007)
	Lower functional connectivity	Task-dependent	(Just et al., 2007)
Ventral occipitotemporal regions	More activation during local processing and visual search	Task-dependent	(Ring et al., 1999)
Prefrontal cortical areas	Less activation during local processing and visual search	Task-dependent	(Ring et al., 1999)
Global functional connectivity	Reduced in adolescents and adults with ASD	Resting state	(Uddin, Supekar, & Menon, 2013)
	Increased in younger children with ASD	Resting state	(Uddin et al., 2013)
	Reduced connectivity within functional systems	Resting state	(Rudie et al., 2013)
	Stronger connectivity between functional systems	Resting state	(Rudie et al., 2013)
Interhemispheric connections	Functional underconnectivity	Resting state	(Anderson et al., 2011)

and long-range connectivity within functional systems and stronger connectivity between functional systems, particularly in default and higher-order visual regions (Rudie et al., 2013). In addition to the differences in functional connectivity seen in ASD, there have been notable functional activation differences in specific areas in ASD and control groups. In resting state fMRI, researchers found that the ASD group had altered functional organization of the resting-state network, which is also referred to as the default mode network (Kennedy & Courchesne, 2008b). The resting-state network is most active when a person is not engaged in a particular task and becomes inactive when a person re-engages with the outside world. Individuals with ASD do not show the same deactivation of the resting-state network that typical controls do, and this lack of deactivation is thought to underlie the lack of social engagement that characterizes the disorder (Kennedy et al., 2006). These differences in functional connectivity and functional activation both explain the deficits seen in ASD and provide rationale for use of functional data in classification of the disorder.

Although functional differences comprise the majority of the differences found between ASD and typical controls (Rudie et al., 2013), researchers have consistently found that individuals with ASD have reduced corpus callosum volume (Anderson et al., 2011; Chen et al., 2011; Pagnozzi, Conti, Calderoni, Fripp, & Rose, 2018) and increased frontal lobe volumes (Brambilla et al., 2003; Chen et al., 2011; Pagnozzi et al., 2018). Beyond these two consistent findings, there is great heterogeneity in the structural differences observed in ASD (Amaral et al., 2008), as shown in Table 4 on the following page. Beyond the varied brain areas reported to be structurally different in ASD, there are multiple studies that directly contradict one another. For example, the cerebellum has been observed to have reduced volume in ASD compared to typical controls (Fatemi et al., 2012; Pagnozzi et al., 2018; Toal et al., 2010), while it has also

Table 4. Functional Differences Between ASD and TDA

Brain Area	Finding	Study
Corpus callosum	Reduced volume	(Anderson et al., 2011; Chen et al., 2011; Pagnozzi et al., 2018)
Whole brain	Increased surface area and increased cortical gyrification	(Pagnozzi et al., 2018)
	Increased CSF volume	(McAlonan et al., 2005; Pagnozzi et al., 2018)
Frontal lobes	Increased GM volume	(Amaral et al., 2008; Chen et al., 2011; Pagnozzi et al., 2018)
	Decreased GM density and abnormal growth trajectories	(Chen et al., 2011)
	Increased cortical thickness	(Pagnozzi et al., 2018)
	Increased WM volume	(Amaral et al., 2008)
Temporal lobes	Increased GM volume	(Brambilla et al., 2003; Chen et al., 2011; Pagnozzi et al., 2018)
	Decreased WM volume and density	(Chen et al., 2011)
	Abnormal growth trajectories	(Chen et al., 2011)
	Decreased GM volume	(McAlonan et al., 2005; Toal et al., 2010)
Parietal lobes	Increased cortical thickness	(Chen et al., 2011)
	Decreased GM	(McAlonan et al., 2005)
	Increased volume	(Brambilla et al., 2003)
Cerebellum	Reduced volume	(Fatemi et al., 2012; Pagnozzi et al., 2018; Toal et al., 2010)
	Increased volume	(Amaral et al., 2008; Brambilla et al., 2003)
	Reduced WM	(McAlonan et al., 2005; Toal et al., 2010)
Fronto-striatal network	Reduced GM	(McAlonan et al., 2005)
Fusiform gyrus	Reduced GM volume	(Toal et al., 2010)
Brainstem	Reduced WM	(Toal et al., 2010)

been found to have increased volume in ASD compared to typical controls (Amaral et al., 2008; Brambilla et al., 2003). While some of the variance in reported structural differences may be due to small numbers of subjects examined within these studies, these findings reinforce the view of autism as a spectrum disorder, arising from a variety of structural and functional differences that underlie the deficits seen in ASD.

Both the importance of earlier diagnosis of ASD and the myriad of structural and functional differences seen in ASD motivate development of tools for diagnosis of ASD that leverage neuroimaging data. Various efforts have been made to use machine learning and neural networks for diagnosis of ASD in recent years. Many of these machine learning approaches are built upon use of MR-derived features such as volume, surface, and thickness. These models have been able to achieve diagnostic accuracy up to 87%, although (1) the data was only from 22 subjects with ASD and 16 controls and (2) the performance of the diagnostic model is critically affected by the features selected as components of the model (Chen et al., 2011; Jiao et al., 2010). Researchers have also used a 3-layer fully connected neural network to classify ASD from controls using 19,900 features derived from fMRI data from 505 individuals with ASD and 530 controls in the Autism Brain Imaging Data Exchange (ABIDE) database, achieving an accuracy of 70%, which is state-of-the-art for the ABIDE dataset (Heinsfeld, Franco, Craddock, Buchweitz, & Meneguzzi, 2018). Classification by the neural network showed an anticorrelation of brain function between the anterior and posterior areas of the brain, corroborating empirical evidence of anterior-posterior disruption in brain connectivity in ASD (Just, Cherkassky, Keller, & Minshew, 2004).

Given the structural and functional differences in ASD and the importance of developing new diagnostic tools that will enable early diagnosis of the disorder, we aim to classify ASD

from structural T1-weighted brain MRI images, 3D volumes of voxel-wise cortical thickness, and mean functional activation volumes using the following classification methods: nearest neighbor (NN), Naïve Bayes, linear discriminant analysis (LDA), and residual neural networks (RNNs). Additionally, we aim is to correlate the distances of imaging volumes from the classification boundary, which is analogous to the classifier's confidence of its prediction, to ADOS scores in order to demonstrate that the classifier mimics classification behavior of the current gold standard assessment tool for ASD. We hypothesize that the machine learning classifiers, as well as the residual neural network, which we call ASDNet, will be able to classify significantly better than chance for all three imaging types. We also hypothesize that the distances of imaging volumes from the classification boundary will be positively correlated with ADOS and ADI-R scores. Implementation of machine learning algorithms, such as neural networks like ASDNet, in combination with careful clinical assessments can potentially change the way ASD is diagnosed, paving the way for faster interventions and better outcomes.

Methods

ABIDE Preprocessed

Autism Brain Imaging Data Exchange (ABIDE) is a collaboration of 16 international imaging sites that have aggregated and are openly sharing structural and resting state functional MRI data along with phenotypic information from 539 individuals with ASD and 573 typical controls (Di Martino et al., 2014). The Preprocessed Connectomes Project (PCP) has publicly released data from ABIDE that was preprocessed by five different teams each using their preferred tools, resulting in a dataset that contains the same raw data preprocessed in a multitude of ways. T_1-weighted structural MRI images from 531 individuals with ASD and 570 typical

controls were downloaded from ABIDE Preprocessed. These structural MRI images were motion-corrected, intensity normalized, Talairach-transformed, and skull-stripped using FreeSurfer prior to being downloaded. The downloaded structural MRI images were reduced to a resolution of 64 x 64 x 64 and min-max normalized in order to make the volumes well-conditioned for training with neural networks. Additionally, 3D volumes of voxel-wise cortical thickness from 531 individuals with ASD and 570 typical controls were utilized. These volumes come from the Advanced Normalization Tools (ANTs) cortical thickness pipeline, which uses multivariate template construction, image registration, bias correction, and *n*-tissue segmentation to perform automated, volume-based estimation of cortical thickness measures from structural MRI data (Avants et al., 2014; Tustison et al., 2014). The ANTs cortical thickness pipeline has good scan-rescan repeatability and demonstrated higher predictive performance than FreeSurfer surface-based cortical thickness estimations for thickness-based prediction of age and gender (Tustison et al., 2014). Following download, these volumes were first cropped to a size of 140 x 120 x 180 in order to eliminate uninformative zeros. All the voxels removed during cropping had a value of zero for all volumes of voxel-wise cortical thickness. After being cropped, the resolution of the volumes was reduced by a factor of two, resulting in a final resolution of 70 x 60 x 90. The cortical thickness volumes were also min-max normalized. Finally, mean functional activation volumes from 529 individuals with ASD and 570 typical controls were used. These volumes were preprocessed using the Connectome Computation System (CCS) pipeline, which integrates tools from AFNI, FSL and FreeSurfer to perform comprehensive functional preprocessing. Following download, min-max normalization was performed on the mean functional activation volumes. The resolution of the volumes was not changed from their original resolution of 61 x 73 x 61.

Classifiers

Following data preprocessing, we used the CoSMoMVPA toolbox (Oosterhof, Connolly, & Haxby, 2016) to classify structural, cortical thickness, and functional MRI scans from individuals with ASD and typically developed controls. We first reduced the dimensionality of each scan to 1100 components using principal components analysis. We then used 8-fold cross validation separately using three classifiers: Nearest neighbor (NN), Naïve Bayes, and linear discriminant analysis (LDA). In this procedure, MRI scans are randomly assigned to one of eight subsets. Seven of the eight subsets (87.5% of the data) are pooled together to train the classifier (NN, Naïve Bayes, LDA) and then classification accuracy is tested on the remaining subset (12.5% of the data). This procedure is repeated a total of eight times, such that each of the subsets is tested once. Classification results are reported in percent correct of classifications for each subset with standard error across subsets. In addition to performing this procedure for structural, cortical thickness, and functional scans independently, we also created ensembles of every combination (i.e. structural + cortical, structural + functional, etc.) of scans for each individual. For the ensembles, we concatenated scans together, such that ensembles with two scan types have 2200 components and ensembles with all three scan types have 3300 components.

Linking ADOS to classification data using distance to bound

To link behavioral metrics from ASD behavioral assessments such as ADI-R and ADOS to cortical, structural, and functional differences in individuals with ASD we used a "distance to bound" analysis, a method that has been used in a number of psychophysics studies (Carlson, Ritchie, Kriegeskorte, Durvasula, & Ma, 2013; Ritchie & Carlson, 2016; Ritchie, Tovar, &

Carlson, 2015). In this analysis, an LDA classifier is used, and the Euclidean distance from the classification boundary of the MRI scan is calculated. A greater distance denotes that the item is more easily categorized into the ASD class type. We then performed a spearman rank correlation between the distance to bound and both ADI-R and ADOS behavioral scores.

ASDNet

For each of the three imaging modalities downloaded, 80% of the volumes were used to train a version of ASDNet specific to that imaging modality. These volumes were presented to the network in batches of three, and the network was modified based on the error of all three volumes presented within the batch. ASDNet saw each of the volumes 172 times during training (each cycle through the training data is called an epoch). At the beginning of each epoch, 10% of the volumes (an equal number from each class) were used to validate how well ASDNet could generalize to data it was not being trained on. The remaining 10% of volumes (also an equal number from each class) were used at the end of training to test the network's ability to classify following training. Five-fold cross-validation was done, resulting in ASDNet being trained five separate times, with no volume serving as part of the validation or test dataset more than once for each of the three imaging modalities.

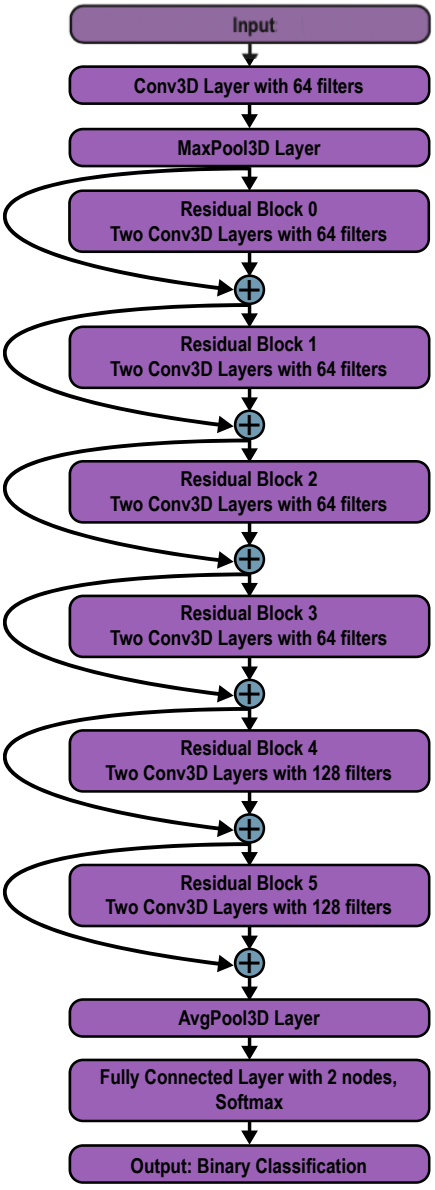

Figure 8. Architecture of ASDNet

ASDNet utilizes the same neural network architecture as the deep 3D residual neural network used to diagnose Alzheimer's disease that achieved a classification accuracy of 79.4% (Yang et al., 2018b). ASDNet consists of 14 layers, 13 of which are convolutional layers and one which is a fully connected layer. The first layer is a convolutional layer, which is followed by max pooling. The output from the max pooling is passed to a series of six residual blocks, each consisting of a shortcut connection from the start of the residual block to its end and two convolutional layers preceded by batch normalization and rectified linear unit (ReLU) activation. The output of the residual blocks is passed to an average pooling operation, and then finally to the fully connected layer, which classifies the structural MRI image as being either ASD or control. The same network architecture was used for neural network classification of T_1-weighted structural MRI images, 3D volumes of voxel-wise cortical thickness, and mean functional activation volumes. The complete architecture of ASDNet is shown in Figure 8 on the previous page.

Results

The classification accuracy for each classifier on the structural, cortical thickness, and function volumes is displayed in Figure 9 on the following page. LDA had the best average classification accuracy across the three types of imaging out of the four classification methods, with ASDNet classifying structural and cortical thickness volumes with similar accuracy and performing worse on the functional volumes than LDA. Naïve-Bayes showed comparable accuracy to LDA and ASDNet on structural MRI, but had performance close to that of nearest neighbors for the cortical thickness and functional volumes. Putting data from two imaging types together into an ensemble increased classification accuracy with all three classifiers it was

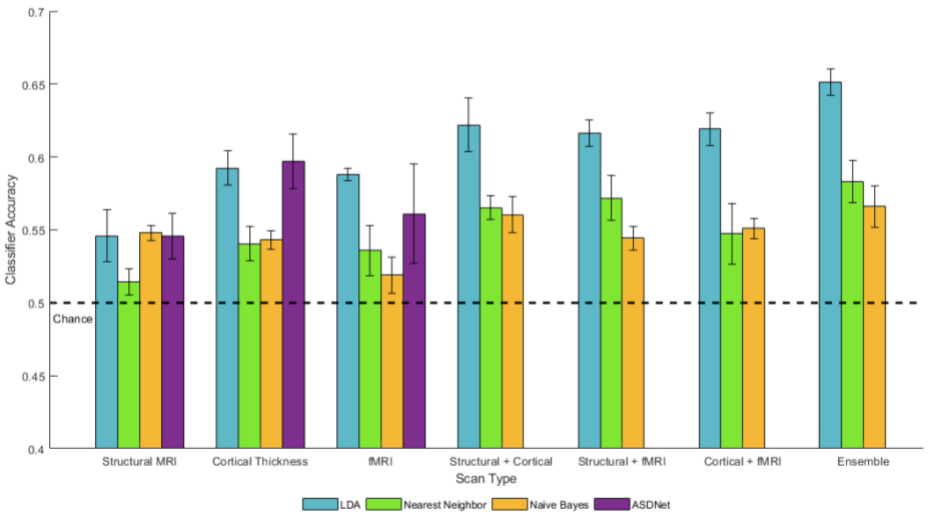

Figure 9.Classification accuracy for each classifier on structural, cortical thickness, and functional volumes.

attempted with (LDA, Naïve-Bayes, and nearest neighbors). The ensemble containing all three imaging types performed the best. Finally, use of cortical thickness volumes and functional volumes led to higher classification than use of structural MRI, although the pairwise combinations of imaging types as an ensemble led to similar classification accuracies for all three pairwise combinations.

Tables 5 & 6 on the following page show the Spearman correlation of distance to bound with ADI-R and ADOS scores. Distance to bound was not significantly correlated with any of the ADI-R scores. The only significant correlation of distance to bound with ADOS scores found with a single imaging type was in the communication subscore for cortical thickness volumes. Distance to bound was significantly correlated with the ADOS total score and the social subscore for the ensemble of structural and functional volumes, while distance to bound was significantly correlated with the ADOS total score and each of the ADOS subscores for the ensemble of cortical thickness and functional MRI volumes. There was also a significant correlation of distance to bound with the ADOS total score and each of the subscores for ensemble of all three imaging types.

Table 5. Spearman Correlation of Distance to Bound with ADI-R Scores

Imaging Type	Reciprocal Social Interaction Subscore	Abnormalities in Communication Subscore	Restricted, Repetitive, and Stereotyped Patterns of Behavior Subscore	Abnormality of Development Evident at or Before 36 Months Subscore
Structural	-0.030824	0.00310786	0.01198704	-0.080839
Cortical Thickness	0.07179363	0.04178351	0.03704429	0.09608714
fMRI	-0.0052388	0.00160912	-0.0462776	0.02620907
Structural + Cortical Thickness	-0.0045388	-0.0546301	-0.0480313	-0.0462883
Structural + fMRI	0.03864535	0.05123317	-0.0201356	-0.0236276
Cortical Thickness +fMRI	0.05882777	0.03533393	-0.031688	-0.0017648
Ensemble	0.05843972	0.01987662	-0.0513222	-0.0096185

* bold denotes $p < 0.05$

Table 6. Spearman Correlation of Distance to Bound with ADOS Scores

Imaging Type	Communication Total Subscore	Social Total Subscore	Stereotyped Behaviors and Restricted Interest Total Subscore	ADOS Total Score
Structural	0.045096734	0.096477124	0.140835223	0.070480834
Cortical Thickness	**0.094878224**	0.102398749	0.114871145	0.087638426
fMRI	0.062210081	0.076896168	-0.03648645	0.075727795
Structural + Cortical Thickness	0.023658304	**0.104344813**	0.149861636	0.083697354
Structural + fMRI	0.083834291	**0.124648891**	0.099444885	**0.113113377**
Cortical Thickness +fMRI	**0.211939743**	**0.239767411**	**0.146099473**	**0.234762654**
Ensemble	**0.140052023**	**0.196146979**	**0.159847403**	**0.185239497**

* bold denotes $p < 0.05$

Discussion

In this study, LDA was the most effective classifier across the three types of imaging, achieving classification accuracies slightly better than ASDNet while taking a fraction of the training time and computation of ASDNet. The fact that classification accuracy was higher for ensembles of two imaging types than for classification with one imaging type alone suggests that the differences between ASD and typical controls in different imaging types can act synergistically to improve classification. An ensemble of multiple imaging types allows the classifier to leverage the differences between ASD and typical controls found in both imaging types, increasing its ability to separate the two groups and accurately predict which of the two groups a test item belongs to. This notion is further corroborated by the ensemble of all three imaging types achieving the highest accuracy seen in this study. The superior performance of the ensembles also indicates that future attempts to diagnose ASD using machine learning and neuroimaging data should make use of data from multiple imaging modalities, as doing so will allow the classifier to use a larger amount of brain differences between ASD and typically developed controls and achieve higher classification. The higher classification accuracy for cortical thickness volumes and functional volumes than for structural volumes intimates that there are more significant differences between ASD and typical controls in cortical thickness and functional MRI than in structural MRI, which is consistent with previous studies that showed structural differences comprise only a small subset of the neuroimaging changes found in ASD (Rudie et al., 2013). This is not to say that differences in structural MRI are not informative for classification of ASD; the ensembles showed that consideration of structural MRI in addition to cortical thickness and/or functional MRI effectively improved classification accuracy.

ASDNet did not perform as well as one would expect given the success deep neural networks have had in medical imaging. While the well-documented heterogeneity of structural and functional changes in ASD may make the problem of ASD classification more difficult than classification of other diseases or disorders, the fact that ASDNet did not significantly outperform the other classifiers indicates that something about the network is lnot functioning optimally. In order to determine whether ASDNet was implemented incorrectly, we created a toy dataset of small and large cubes with the same resolution as the structural MRI images. ASDNet learned how to classify the test set correctly within 20 training steps (i.e. 20 iterations of batch presentation followed by weight modification through backpropagation), supporting the notion that ASDNet was implemented correctly and suggesting that the network is capable of solving high-dimensional classification problems. To further demonstrate that the implementation is correct, ASDNet should be trained and tested on the same Alzheimer's Disease Neuroimaging Initiative (ADNI) data that was used to achieve a classification accuracy of 79.4% by Yang et al. (2018b). Replication of these findings would confirm that ASDNet has been implemented as intended.

Another possible reason for the suboptimal performance of ASDNet is that the parameters that dictate how ASDNet is trained, which are called hyperparameters, are flawed. Deep convolutional neural networks like ASDNet have around ten to fifty hyperparameters, many of which are continuous values, so the number of ways in which the values of these hyperparameters can be permuted is huge. Given that (1) these neural networks often take hours or days to train and (2) the particular set of hyperparameters can drastically affect the efficacy of training, selection of an effective configuration of hyperparameters can be both time-consuming and difficult (Bergstra, Bardenet, Bengio, & Kégl, 2011). Bayesian optimization has been shown

to obtain better results in fewer evaluations than other optimization algorithms like grid search and random search because Bayesian optimization is able to reason about the quality of a hyperparameter configuration before it is run (Snoek, Larochelle, & Adams, 2012). We have tried using Bayesian optimization to find a good set of hyperparameters and the "optimal" hyperparameters identified by this optimization were used to train ASDNet. Because the optimization algorithm uses the performance of past hyperparameter configurations to determine which set of hyperparameters to use next, one cannot perform more than one run at a time during optimization. This inability to parallelize Bayesian optimization, coupled with the fact that ASDNet takes two to six hours per run, severely limited both the number of runs that could be done within an instance of optimization and the number of hyperparameters that really could be optimized. Bayesian optimization of a larger number of hyperparameters for a longer amount of time may enable the optimization algorithm to thoroughly explore the range of possible hyperparameter configurations and find a truly effective set of hyperparameters.

The distance to bound analysis demonstrated that structural, cortical thickness, and functional differences extracted by the classifier for classification of ASD are linked to metrics from behavioral assessments. The fact that distance to bound was significantly correlated with ADOS scores but not with any of the ADI-R scores suggests that ADOS scores account for the structural, cortical thickness, and functional differences better than ADI-R, which is consistent with findings that ADOS is more sensitive to ASD than ADI-R (Randall et al., 2018). There was greater and more significant correlation of distance to bound with ADOS scores when data from multiple imaging types were grouped into an ensemble, reinforcing the conclusion from the classification accuracy results that providing the classifier with differences from multiple imaging types increased its ability to separate the two groups. Moreover, the significant

correlation of distance to bound with the ADOS total score and each of the subscores for the ensemble of cortical thickness and functional MRI intimates that the cortical thickness and functional differences in ASD may underlie the range of behavioral deficits seen in ASD that ADOS tests for. Additionally, the higher, more significant correlation for the ensemble of cortical thickness and functional MRI than for either of the ensembles involving structural MRI further strengthens the conclusion from the results of classifier performance that the cortical thickness and functional differences likely contribute more to ASD than structural differences do. Thus, not only did the distance to bound analysis demonstrate that brain differences in ASD extracted by the classifier are linked to metrics from behavioral assessments, it reinforced many of the conclusions drawn from the results of classifier performance.

Furthermore, the classifiers explored in this study have the potential to enable earlier diagnosis and thus earlier treatment of ASD. Current behavioral assessments can be only used with children 12 months and older (Esler, Amy N; Bal, Vanessa H; Guthrie, Whitney; Weismer, Susan E; Lord, 2015), while classification of ASD using machine learning could be done in children under 12 months of age. A classifier that diagnoses ASD for children under 12 months of age would be constructed and trained in the same manner that classifiers in this study have been, the only major difference being that the classifier would need to be trained and tested on neuroimaging scans from children under 12 months of age. Such classifiers would make early diagnosis and early intervention more feasible and provide immediate health benefits. Thus, this study demonstrates that machine learning may be a viable way to improve upon current behavioral assessments, assess age groups that were previously too young to be assessed with behavioral exams, and lower healthcare costs by enabling early, accurate diagnoses.

References

Acharya, U. R., Fujita, H., Lih, O. S., Adam, M., Tan, J. H., & Chua, C. K. (2017). Automated detection of coronary artery disease using different durations of ECG segments with convolutional neural network. *Knowledge-Based Systems*, *132*, 62–71. https://doi.org/10.1016/j.knosys.2017.06.003

Acharya, U. R., Fujita, H., Oh, S. L., Hagiwara, Y., Tan, J. H., & Adam, M. (2017). Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. *Information Sciences*, *415–416*, 190–198. https://doi.org/10.1016/j.ins.2017.06.027

Acharya, U. R., Oh, S. L., Hagiwara, Y., Tan, J. H., & Adeli, H. (2018). Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. *Computers in Biology and Medicine*, *100*(July 2017), 270–278. https://doi.org/10.1016/j.compbiomed.2017.09.017

Amaral, D. G., Schumann, C. M., & Nordahl, C. W. (2008). Neuroanatomy of autism. *Trends in Neurosciences*, *31*(3), 137–145. https://doi.org/10.1016/j.tins.2007.12.005

Anderson, J. S., Druzgal, T. J., Froehlich, A., Dubray, M. B., Lange, N., Alexander, A. L., … Lainhart, J. E. (2011). Decreased interhemispheric functional connectivity in autism. *Cerebral Cortex*, *21*(5), 1134–1146. https://doi.org/10.1093/cercor/bhq190

Anthimopoulos, M., Christodoulidis, S., Ebner, L., Christe, A., & Mougiakakou, S. (2016). Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network. *IEEE Transactions on Medical Imaging*, *35*(5), 1207–1216. https://doi.org/10.1109/TMI.2016.2535865

Avants, B. B., Tustison, N. J., Stauffer, M., Song, G., Wu, B., & Gee, J. C. (2014). The Insight ToolKit image registration framework. *Frontiers in Neuroinformatics*, *8*(April), 1–13. https://doi.org/10.3389/fninf.2014.00044

Battaglia, A., & Carey, J. C. (2006). Etiologic Yield of Autistic Spectrum Disorders: A Prospective Study. *American Journal of Medical Genetics*, (142C), 3–7. https://doi.org/10.1002/ajmg.c

Bergstra, J., Bardenet, R., Bengio, Y., & Kégl, B. (2011). Algorithms for Hyper-Parameter Optimization. *NIPS*, 1–9. https://doi.org/2012arXiv1206.2944S

Bernal, J., Kushibar, K., Asfaw, D. S., Valverde, S., Oliver, A., Martí, R., & Lladó, X. (2017). Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: a review. Retrieved from http://arxiv.org/abs/1712.03747

Bourgeron, T. (2015). From the genetic architecture to synaptic plasticity in autism spectrum disorder. *Nature Reviews Neuroscience*, *16*(9), 551–563. https://doi.org/10.1038/nrn3992

Brambilla, P., Hardan, A., Ucelli Di Nemi, S., Perez, J., Soares, J. C., & Barale, F. (2003). Brain anatomy and development in autism: Review of structural MRI studies. *Brain Research Bulletin*, *61*(6), 557–569. https://doi.org/10.1016/j.brainresbull.2003.06.001

Brentani, H., de Paula, C. S., Bordini, D., Rolim, D., Sato, F., Portolese, J., … McCracken, J. T. (2013). Autism spectrum disorders: An overview on diagnosis and treatment. *Revista Brasileira de Psiquiatria*, *35*(SUPPL. 1), 62–72. https://doi.org/10.1590/1516-4446-2013-S104

Büyüköztürk, Ş., & Çokluk-Bökeoğlu, Ö. (2008). Discriminant function analysis: Concept and application. *Egitim Arastirmalari - Eurasian Journal of Educational Research*, (33), 73–92.

Carlson, T. A., Ritchie, J. B., Kriegeskorte, N., Durvasula, S., & Ma, J. (2013). Reaction Time

for Object Categorization Is Predicted by Representational Distance. *Journal of Cognitive Neuroscience*, *26*(1), 132–142. https://doi.org/10.1162/jocn

Chasson, G. S., Harris, G. E., & Neely, W. J. (2007). Cost comparison of early intensive behavioral intervention and special education for children with autism. *Journal of Child and Family Studies*, *16*(3), 401–413. https://doi.org/10.1007/s10826-006-9094-1

Chen, R., Jiao, Y., & Herskovits H., E. (2011). Structural MRI in autism spectrum disorder. *Pediatric Research*, *69*, 1–13. https://doi.org/10.1203/PDR.0b013e318212c2b3.Structural

Cherkassky, V. L., Kana, R. K., Keller, T. A., & Just, M. A. (2006). Functional connectivity in a baseline resting state network in autism. *NeuroReport*, *17*(16), 7–10. Retrieved from /paper/Functional-connectivity-in-a-baseline-resting-Cherkassky-Kana/002dcf48cbdc58bab4aa6a3b90e33575bf4c353a

Coomans, D., & Massart, D. L. (1982). Alternative k-nearest neighbour rules in supervised pattern recognition. *Analytica Chimica Acta*, *136*, 15–27. https://doi.org/10.1016/s0003-2670(01)85298-3

Cooper, A. S., Friedlaender, E., Levy, S. E., Shekdar, K. V., Bradford, A. B., Wells, K. E., & Mollen, C. (2016). The Implications of Brain MRI in Autism Spectrum Disorder. *Journal of Child Neurology*, *31*(14), 1611–1616. https://doi.org/10.1177/0883073816665548

Courchesne, E., Campbell, K., & Solso, S. (2011). Brain growth across the life span in autism: Age-specific changes in anatomical pathology. *Brain Research*, *1380*, 138–145. https://doi.org/10.1016/j.brainres.2010.09.101

Das, R., Turkoglu, I., & Sengur, A. (2009). Effective diagnosis of heart disease through neural networks ensembles. *Expert Systems with Applications*, *36*(4), 7675–7680. https://doi.org/10.1016/j.eswa.2008.09.013

Di Martino, A., Yan, C. G., Li, Q., Denio, E., Castellanos, F. X., Alaerts, K., … Milham, M. P. (2014). The autism brain imaging data exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular Psychiatry*, *19*(6), 659–667. https://doi.org/10.1038/mp.2013.78

Elder, J. H., Kreider, C. M., Brasher, S. N., & Ansell, M. (2017). Clinical impact of early diagnosis of autism on the prognosis and parent-child relationships. *Psychology Research and Behavior Management*, *10*, 283–292. https://doi.org/10.2147/PRBM.S117499

Emerson, R. W., Adams, C., Nishino, T., Hazlett, H. C., Wolff, J. J., Zwaigenbaum, L., … Piven, J. (2017). Functional neuroimaging of high-risk 6-month-old infants predicts a diagnosis of autism at 24 months of age. *Science Translational Medicine*, *9*(393). https://doi.org/10.1126/scitranslmed.aag2882.Functional

Esler, Amy N; Bal, Vanessa H; Guthrie, Whitney; Weismer, Susan E; Lord, C. (2015). The Autism Diagnostic Observation Schedule, Toddler Module: Standardized Severity Scores. *J Autism Dev Disord*, *91*(2), 165–171. https://doi.org/10.1016/j.chemosphere.2012.12.037.Reactivity

Falkmer, T., Anderson, K., Falkmer, M., & Horlin, C. (2013). Diagnostic procedures in autism spectrum disorders: A systematic literature review. *European Child and Adolescent Psychiatry*, *22*(6), 329–340. https://doi.org/10.1007/s00787-013-0375-0

Fatemi, S. H., Aldinger, K. A., Blaha, C. D., Blatt, G. J., Chauhan, A., Kemper, T. L., … Welsh, J. P. (2012). Consensus Paper : Pathological Role of the Cerebellum in Autism. *The Cerebellum*, 777–807. https://doi.org/10.1007/s12311-012-0355-9

Fong, R. C., & Vedaldi, A. (2017). Interpretable explanations of black boxes by meaningful perturbation. *Proceedings of the 2017 IEEE International Conference on Computer Vision*

(ICCV). Retrieved from http://openaccess.thecvf.com/content_ICCV_2017/papers/Fong_Interpretable_Explanations_of_ICCV_2017_paper.pdf

Frith, U., & Happé, F. (2005). Autism spectrum disorder. *Current Biology*, *15*(19), 786–790. https://doi.org/10.1016/j.cub.2005.09.033

Fukushima, K. (1980). Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position. *Biological Cybernetics*, *202*. https://doi.org/10.1007/BF00344251

Grira, N., Crucianu, M., & Boujemaa, N. (2004). Unsupervised and Semi-supervised Clustering: A Brief Survey. *A Review of Machine Learning Techniques for Processing Multimedia Content, Report of the MUSCLE European Network of Excellence*, 1–12. https://doi.org/10.1021/jp303679y

Hand, D. J., & Yu, K. (2001). Idiot's Bayes - Not so stupid after all? *International Statistical Review*, *69*(3), 385–398. https://doi.org/10.1111/j.1751-5823.2001.tb00465.x

Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., … Larochelle, H. (2017). Brain tumor segmentation with Deep Neural Networks. *Medical Image Analysis*, *35*, 18–31. https://doi.org/10.1016/j.media.2016.05.004

Hazlett, H. C., Gu, H., Munsell, B. C., Kim, S. H., Styner, M., Wolff, J. J., … Piven, J. (2017). Early brain development in infants at high risk for autism spectrum disorder. *Nature*, *542*(7641), 348–351. https://doi.org/10.1038/nature21369

He, K., Zhang, X., Ren, S., & Sun, J. (2015a). Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1–17. https://doi.org/10.1007/s11042-017-4440-4

He, K., Zhang, X., Ren, S., & Sun, J. (2015b). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *Proceedings of the IEEE International Conference on Computer Vision*, *2015 Inter*, 1026–1034. https://doi.org/10.1109/ICCV.2015.123

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Identity mappings in deep residual networks. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9908 LNCS*, 630–645. https://doi.org/10.1007/978-3-319-46493-0_38

Heinsfeld, A. S., Franco, A. R., Craddock, R. C., Buchweitz, A., & Meneguzzi, F. (2018). Identification of autism spectrum disorder using deep learning and the ABIDE dataset. *NeuroImage: Clinical*, *17*(June 2017), 16–23. https://doi.org/10.1016/j.nicl.2017.08.017

Herman, G. T., De Pierro, A. R., & Gai, N. (1992). On methods for maximum a posteriori image reconstruction with a normal prior. *Journal of Visual Communication and Image Representation*, *3*(4), 316–324. https://doi.org/10.1016/1047-3203(92)90035-R

Hopfield, J. J. (1988). Artificial Neural Networks. *IEEE Circuits and Devices Magazine*.

Ioffe, S., & Szegedy, C. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. https://doi.org/10.1007/s13398-014-0173-7.2

Jiao, Y., Chen, R., Ke, X., Chu, K., Lu, Z., & Herskovits, E. H. (2010). Predictive models of autism spectrum disorder based on brain regional cortical thickness. *NeuroImage*, *50*(2), 589–599. https://doi.org/10.1016/j.neuroimage.2009.12.047

John, G. H., & Langley, P. (1995). Estimating Continuous Distributions in Bayesian Classifiers. *Proceedings of the Eleventh Conference on Uncertainty in Artifical Intelligence*.

Just, M. A., Cherkassky, V. L., Keller, T. A., Kana, R. K., & Minshew, N. J. (2007). Functional

and anatomical cortical underconnectivity in autism: Evidence from an fmri study of an executive function task and corpus callosum morphometry. *Cerebral Cortex*, *17*(4), 951–961. https://doi.org/10.1093/cercor/bhl006

Just, M. A., Cherkassky, V. L., Keller, T. A., & Minshew, N. J. (2004). Cortical activation and synchronization during sentence comprehension in high-functioning autism: Evidence of underconnectivity. *Brain*, *127*(8), 1811–1821. https://doi.org/10.1093/brain/awh199

Kennedy, D. P., & Courchesne, E. (2008a). Functional abnormalities of the default network during self- and other-reflection in autism. *Social Cognitive and Affective Neuroscience*, *3*(2), 177–190. https://doi.org/10.1093/scan/nsn011

Kennedy, D. P., & Courchesne, E. (2008b). The intrinsic functional organization of the brain is altered in autism. *NeuroImage*, *39*(4), 1877–1885. https://doi.org/10.1016/j.neuroimage.2007.10.052

Kennedy, D. P., Redcay, E., & Courchesne, E. (2006). Failing to deactivate: Resting functional abnormalities in autism. *Proceedings of the National Academy of Sciences*, *103*(21), 8275–8280. https://doi.org/10.1073/pnas.0600674103

Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2006). Machine learning: A review of classification and combining techniques. *Artificial Intelligence Review*, *26*(3), 159–190. https://doi.org/10.1007/s10462-007-9052-3

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances In Neural Information Processing Systems*, 1–9. https://doi.org/http://dx.doi.org/10.1016/j.protcy.2014.09.007

Lachenbruch, P. A., & Goldstein, M. (1979). Discriminant Analysis. *Biometrics*, *35*(1), 69–85.

Lai, M. (2015). Deep Learning for Medical Image Segmentation. Retrieved from http://arxiv.org/abs/1505.02000

LeCun, Y., Haffner, P., Bottou, L., & Bengio, Y. (1999). Object Recognition with Gradient-Based Learning. *Shape, Contour and Grouping in Computer Vision*, 319–345.

Lenroot, R. K., & Yeung, P. K. (2013). Heterogeneity within Autism Spectrum Disorders: What have We Learned from Neuroimaging Studies? *Frontiers in Human Neuroscience*, *7*(October), 1–16. https://doi.org/10.3389/fnhum.2013.00733

Lewis, J. D., Evans, A. C., Jr, J. R. P., Botteron, K. N., McKinstry, R. C., Zwaigenbaum, L., … Piven, J. (2017). The emergence of network inefficiencies in infants with autism spectrum disorder. *Biological Psychiatry*, *82*(3), 176–181. https://doi.org/10.1016/j.otsr.2015.07.012

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., Dilavore, P. C., … Rutter, M. (2000). The Autism Diagnostic Observation Schedule–Generic: A Standard Measure of Social and Communication Deficits Associated with the Spectrum of Autism. *Journal of Autism and Developmental Disorders*, *30*(3). https://doi.org/10.1023/A:1005592401947

Lord, C., Rutter, M., & Couteur, A. Le. (1994). Autism Diagnostic Interview-Revised: A Revised Version of a Diagnostic Interview for Caregivers of Individuals with Possible Pervasive Developmental Disorders. *Journal of Autism and Developmental Disorders*, *24*(5). https://doi.org/10.1007/BF02172145

Luyster, R., Gotham, K., Guthrie, W., Coffing, M., Petrak, R., Pierce, K., … Lord, C. (2009). The Autism Diagnostic Observation Schedule – Toddler Module: A new module of a standardized diagnostic measure for autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *39*(9), 1305–1320. https://doi.org/10.1007/s10803-009-0746-z.The

Maroco, J., Silva, D., Rodrigues, A., Guerreiro, M., Santana, I., & De Mendonça, A. (2011).

Data mining methods in the prediction of Dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests. *BMC Research Notes*, *4*. https://doi.org/10.1186/1756-0500-4-299

McAlonan, G. M., Cheung, V., Cheung, C., Suckling, J., Lam, G. Y., Tai, K. S., … Chua, S. E. (2005). Mapping the brain in autism. A voxel-based MRI study of volumetric differences and intercorrelations in autism. *Brain*, *128*(2), 268–276. https://doi.org/10.1093/brain/awh332

Muhle, R. A., Reed, H. E., Stratigos, K. A., & Veenstra-VanderWeele, J. (2018). The emerging clinical neuroscience of autism spectrum disorder a review. *JAMA Psychiatry*, *75*(5), 514–523. https://doi.org/10.1001/jamapsychiatry.2017.4685

Myers, S. M., & Johnson, C. P. (2007). Management of Children With Autism Spectrum Disorders. *Pediatrics*, *120*(5), 1162–1182. https://doi.org/10.1542/peds.2007-2362

Nair, V., & Hinton, G. E. (2010). Rectified Linear Units Improve Restricted Boltzmann Machines. *ICML*. https://doi.org/10.1.1.165.6419

Oh, S. L., Hagiwara, Y., Raghavendra, U., Yuvaraj, R., Arunkumar, N., Murugappan, M., & Acharya, U. R. (2018). A deep learning approach for Parkinson's disease diagnosis from EEG signals. *Neural Computing and Applications*, *5*. https://doi.org/10.1007/s00521-018-3689-5

Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (2016). CoSMoMVPA: Multi-Modal Multivariate Pattern Analysis of Neuroimaging Data in Matlab/GNU Octave. *Frontiers in Neuroinformatics*, *10*(July), 1–27. https://doi.org/10.3389/fninf.2016.00027

Ozyilmaz, L., & Yildirim, T. (2002). Diagnosis of thyroid disease using artificial neural network methods. *ICONIP 2002 - Proceedings of the 9th International Conference on Neural Information Processing: Computational Intelligence for the E-Age*, *4*, 2033–2036. https://doi.org/10.1109/ICONIP.2002.1199031

Pagnozzi, A. M., Conti, E., Calderoni, S., Fripp, J., & Rose, S. E. (2018). A systematic review of structural MRI biomarkers in autism spectrum disorder: A machine learning perspective. *International Journal of Developmental Neuroscience*, *71*(April), 68–82. https://doi.org/10.1016/j.ijdevneu.2018.08.010

Payan, A., & Montana, G. (2015). Predicting Alzheimer's disease: a neuroimaging study with 3D convolutional neural networks, 1–9. https://doi.org/10.1613/jair.301

Pereira, S., Pinto, A., Alves, V., & Silva, C. A. (2016). Brain Tumor Segmentation using Convolutional Neural Networks in MRI Images. *IEEE Transactions on Medical Imaging*, *35*(5), 1240–1251. https://doi.org/10.1109/TMI.2016.2538465

Randall, M., Egberts, K., Samtani, A., Scholtten, R., Hooft, L., Livingston, N., … Williams, K. (2018). Diagnostic tests for Autism Spectrum Disorders (ASD) in preschool children. *Cochrane Database of Systematic Reviews*, (7). https://doi.org/10.1002/14651858.cd009044

Redcay, E., & Courchesne, E. (2005). When is the brain enlarged in autism? A meta-analysis of all brain size reports. *Biological Psychiatry*, *58*(1), 1–9. https://doi.org/10.1016/j.biopsych.2005.03.026

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?" Explaining the Predictions of Any Classifier. https://doi.org/10.1145/2939672.2939778

Ring, H. A., Baron-Cohen, S., Wheelwright, S., Williams, S. C. R., Brammer, M., Andrew, C., & Bullmore, E. T. (1999). Cerebral correlates of preserved cognitive skills in autism. *Brain*, *122*(7), 1305–1315. https://doi.org/10.1093/brain/122.7.1305

Rish, I. (2001). An empirical study of the naive Bayes classifier, 41–46.
Ritchie, J. B., & Carlson, T. A. (2016). Neural decoding and "inner" psychophysics: A distance-to-bound approach for linking mind, brain, and behavior. *Frontiers in Neuroscience*, *10*(APR), 1–8. https://doi.org/10.3389/fnins.2016.00190
Ritchie, J. B., Tovar, D. A., & Carlson, T. A. (2015). Emerging Object Representations in the Visual System Predict Reaction Times for Categorization. *PLoS Computational Biology*, *11*(6). https://doi.org/10.1371/journal.pcbi.1004316
Rudie, J. D., Brown, J. A., Beck-pancer, D., Hernandez, L. M., Dennis, E. L., Thompson, P. M., … Dapretto, M. (2013). Altered functional and structural brain network organization in autism. *NeuroImage*, *2*, 79–94. https://doi.org/10.1016/j.nicl.2012.11.006
Rumelhart, D. E., Hinton, G. E., & McClelland, J. L. (1986). A General Framework for Parallel Distributed Processing. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition - Volume 1* (pp. 45–76).
Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Proceedings of the IEEE International Conference on Computer Vision*, *2017–Octob*, 618–626. https://doi.org/10.1109/ICCV.2017.74
Shen, M. D., Nordahl, C. W., Young, G. S., Wootton-Gorges, S. L., Lee, A., Liston, S. E., … Amaral, D. G. (2013). Early brain enlargement and elevated extra-axial fluid in infants who develop autism spectrum disorder. *Brain*, *136*(9), 2825–2835. https://doi.org/10.1093/brain/awt166
Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition, 1–14. https://doi.org/10.1016/j.infsof.2008.09.005
Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical Bayesian Optimization of Machine Learning Algorithms. *NIPS*, 1–9. Retrieved from http://arxiv.org/abs/1206.2944
Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., … Arbor, A. (2014). Going Deeper with Convolutions, 1–9. https://doi.org/10.1109/CVPR.2015.7298594
Toal, F., Daly, E. M., Page, L., Deeley, Q., Hallahan, B., Bloemen, O., … Murphy, D. G. M. (2010). Clinical and anatomical heterogeneity in autistic spectrum disorder: a structural MRI study. *Psychological Medicine*, *40*(07), 1171–1181. https://doi.org/10.1017/S0033291709991541
Tustison, N. J., Cook, P. A., Klein, A., Song, G., Das, S. R., Duda, J. T., … Avants, B. B. (2014). Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements. *Neuroimage*.
Uddin, L. Q., Supekar, K., & Menon, V. (2013). Reconceptualizing functional brain connectivity in autism from a developmental perspective. *Frontiers in Human Neuroscience*, *7*(August), 1–11. https://doi.org/10.3389/fnhum.2013.00458
Williams, D. (2007). Understanding Autism and Related Disorders: What Has Imaging Taught Us? *Neuroimaging Clin N Am.*, *17*(4), 495–ix.
Wolff, J. J., Jacob, S., & Elison, J. T. (2018). The journey to autism: Insights from neuroimaging studies of infants and toddlers. *Development and Psychopathology*, *30*(2), 479–495. https://doi.org/10.1038/nrg3575.Systems
Wolff, J. J., Ph, D., Gu, H., Gerig, G., Elison, J. T., Dawson, G., … Zwaigenbaum, L. (2012). Differences in White Matter Fiber Tract Development Present from 6 to 24 Months in Infants with Autism. *American Journal of Psychiatry*, *169*(6), 589–600. https://doi.org/10.1176/appi.ajp.2011.11091447

Yang, C., Rangarajan, A., & Ranka, S. (2018a). Global Model Interpretation via Recursive Partitioning. *The 4th IEEE International Conference on Data Science and Systems*. https://doi.org/10.1103/PhysRevB.80.081304

Yang, C., Rangarajan, A., & Ranka, S. (2018b). Visual Explanations From Deep 3D Convolutional Neural Networks for Alzheimer's Disease Classification. Retrieved from http://arxiv.org/abs/1803.02544

Zhang, Z. (2016). Introduction to machine learning: k-nearest neighbors. *Annals of Translational Medicine*, *4*(11), 218–218. https://doi.org/10.21037/atm.2016.03.37

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning Deep Features for Discriminative Localization. *The IEEE Conference on Computer Vision and Pattern Recognition*, 2921–2929. https://doi.org/10.1109/CVPR.2016.319