

**ASD PREDICTION FROM STRUCTURAL MRI WITH MACHINE  
LEARNING**

by

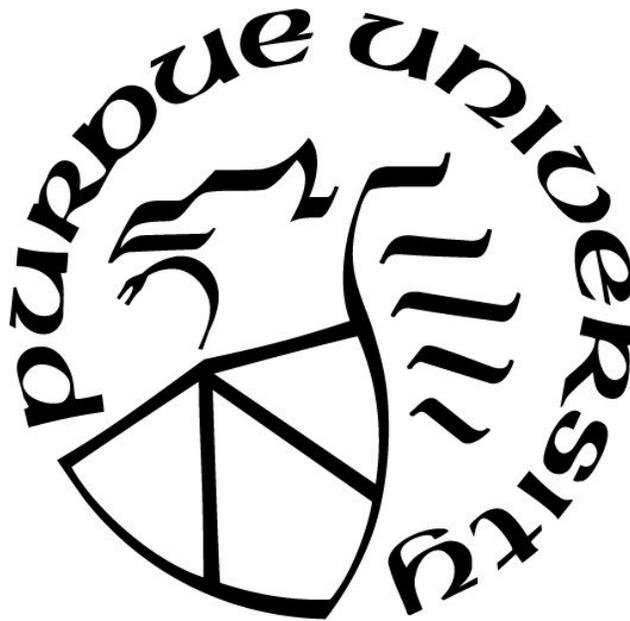
**Nanxin Jin**

**A Thesis**

*Submitted to the Faculty of Purdue University*

*In Partial Fulfillment of the Requirements for the Degree of*

**Master of Science**



Department of Computer and Information Technology

West Lafayette, Indiana

May 2020

**THE PURDUE UNIVERSITY GRADUATE SCHOOL**  
**STATEMENT OF COMMITTEE APPROVAL**

Dr. Baijian Yang, Chair

Department of Computer and Information Technology

Dr. Tonglin Zhang

Department of Statistics

Dr. Byung-Cheol Min

Department of Computer and Information Technology

**Approved by:**

Dr. Eric T. Matson

Head of the Graduate Program

Dedicated to all medical staff, soldiers, police officers, and all first responders who are fighting  
COVID-19.

## **ACKNOWLEDGMENTS**

First, I wish to thank my wife Jiawen Yao for supporting and loving me. My wife always encourages me when I face challenges and takes care of me when I get home. Especially, thank my wife for bringing our upcoming baby to our family.

Next, I would like to acknowledge my parents Yan Jin, Wenjie Sun for supporting me over the past 26 years. I could not study abroad and achieve my goal without their help.

Then, I am really grateful for my advisor Dr. Baijian Yang. Dr. Yang has been helping me on my first day in Purdue graduate school. Dr. Yang not only provided valuable feedback for my research, but also gave me meaningful advice for my career. I really appreciate Dr. Yang for helping me accomplish my master study.

Finally, I wish to gratefully acknowledge my thesis committee Dr. Tonglin Zhang, Dr. Byung-Cheol Min and my research partner Ziyang Tang for their in-depth comments and invaluable suggestions.

# TABLE OF CONTENTS

LIST OF TABLES . . . . .	7
LIST OF FIGURES . . . . .	8
LIST OF ABBREVIATIONS . . . . .	9
GLOSSARY . . . . .	10
ABSTRACT . . . . .	11
CHAPTER 1. INTRODUCTION . . . . .	12
1.1 Background . . . . .	12
1.2 Problem Statement . . . . .	13
1.3 Scope of the Problem . . . . .	15
1.4 Significance of the Problem . . . . .	16
1.5 Statement of Purpose . . . . .	17
1.6 Research Questions . . . . .	17
1.7 Assumptions . . . . .	17
1.8 Limitations . . . . .	18
1.9 Delimitation . . . . .	18
1.10 Summary . . . . .	19
CHAPTER 2. REVIEW OF LITERATURE . . . . .	20
2.1 Overview of Autism Spectrum Disorder . . . . .	20
2.1.1 ASD Incentive . . . . .	21
2.1.2 The Differences of Grey Matter . . . . .	22
2.1.3 ASD Diagnosis . . . . .	23
2.2 Overview of Different Machine Learning Models . . . . .	25
2.2.1 Convolutional Neural Network Overview . . . . .	26
2.2.1.1 CNN Structure . . . . .	27
2.2.1.2 Conv3D: 3 dimensional CNN . . . . .	29
2.2.2 Visual Geometry Group . . . . .	30
2.2.3 Residual Networks . . . . .	32
2.3 Current Study of Applying ML to ASD Diagnosis . . . . .	34

2.4	Summary . . . . .	35
CHAPTER 3. METHODOLOGY . . . . .		36
3.1	Hypothesis . . . . .	36
3.2	Experiment Environment . . . . .	36
3.2.1	Hardware . . . . .	37
3.2.2	Software . . . . .	37
3.3	Data . . . . .	38
3.3.1	Data Source and Anonymization . . . . .	38
3.3.2	Data Description . . . . .	39
3.4	ML Model and Algorithms . . . . .	41
3.4.1	Conv3D . . . . .	42
3.4.2	VGG-16 . . . . .	45
3.4.3	3D ResNet . . . . .	46
3.4.4	Parameters Configuration . . . . .	48
3.5	Variables . . . . .	49
3.6	Testing Procedures . . . . .	50
3.7	Analysis . . . . .	51
3.8	Summary . . . . .	51
CHAPTER 4. RESULTS . . . . .		52
4.1	Conv3D Results . . . . .	52
4.2	VGG Results . . . . .	53
4.3	3D ResNet Results . . . . .	53
4.4	Results Comparison Among Different Models . . . . .	53
4.5	Results Compared with Behavioral Screening Tools . . . . .	55
4.6	Discussion . . . . .	57
4.7	Summary . . . . .	57
CHAPTER 5. CONCLUSION . . . . .		58
5.1	Future Works . . . . .	59
REFERENCES . . . . .		60

## LIST OF TABLES

2.1	Some famous models in ILSVR . . . . .	26
2.2	Different studies of applying ML to ASD prediction. . . . .	35
3.1	The first hardware setup. . . . .	37
3.2	The second hardware setup. . . . .	37
3.3	The software specification. . . . .	38
3.4	Imported ML packages. . . . .	38
3.5	The statistics of dataset. . . . .	39
3.6	The Conv3D model. . . . .	44
3.7	The 3D-VGG model. . . . .	46
3.8	The 3D ResNet model. . . . .	48
4.1	The confusion matrix. . . . .	52
4.2	The best result for Conv3D. . . . .	52
4.3	The best result for VGG. . . . .	53
4.4	The best result for 3D ResNet. . . . .	53

## LIST OF FIGURES

2.1	Grey Matter is located between the green line and red line. The green line represents the pial-grey matter boundary, and the red line represents grey matter-white matter boundary. . . . .	22
2.2	The workflow of DSM-IV for ASD diagnosis (Centers for Disease Control and Prevention., 2020b). . . . .	24
2.3	The workflow of ML method for ASD diagnosis. . . . .	24
2.4	Different number of layers used in different models. . . . .	27
2.5	Convolutional Neural Network structure (Albelwi & Mahmood, 2017). . . . .	27
2.6	(a)2D convolution with kernel size $k*k$ (b)2D convolution on L-frames with kernel size $k*k$ (c)3D convolution on L-frame with kernel size $k*k*d$ . . . . .	29
2.7	Demonstration of how to use two $3*3$ kernel to replace one $5*5$ kernel. . . . .	30
2.8	VGG network structure. Column D represents the structure of VGG-16. Column E represents the structure of VGG-19. . . . .	31
2.9	A building block structure from ResNet (He, Zhang, Ren, & Sun, 2016). . . . .	32
2.10	Left: the regular building block; Right: the "bottleneck" block (He et al., 2016). . .	33
3.1	On the top row, images show the 3rd, 32th and 50th slice of a ASD patient brain MRI. On the bottom row, images show the 80th, 110th and 156th slice of the same ASD patient brain MRI. . . . .	40
3.2	The grey box shows the cropping boundary. The coordinator for the top left vertex is $(64, 64)$ . . . . .	41
3.3	There are two $(3*3*3, 64)$ kernels in each block. . . . .	47
4.1	Models' Performance Comparison. . . . .	54
4.2	Models' Speed Comparison. . . . .	55
4.3	Accuracy Comparison: ML Method vs. Behavioral Test . . . . .	56
4.4	Speed Comparison:ML Method vs. Behavioral Test Speed . . . . .	56

## **LIST OF ABBREVIATIONS**

ADI-R	Autism Diagnostic Interview–Revised
ADOS	The Autism Diagnostic Observation Schedule
DSM-IV	Diagnostic and Statistical Manual of Mental Disorders
ASD	Autism Spectrum Disorders
CDC	Centers for Disease Control and Prevention
CIT	Computer and Information Technology
CNN	Convolutional Neural Networks
CoT	College of Technology
ML	Machine Learning
MRI	Magnetic Resonance Imaging
ResNet	Residual Neural Network
TD	Typical Development

## **GLOSSARY**

Classifier – A classifier is commonly used in Machine Learning and statistical study to identify which category a new observation belongs to.

Freesufer – Freesufer is a software that is able to reconstruct and visualize the human brain(Fischl, 2012).

Magnetic Resonance Imaging – A medical technique to visualize anatomical structures of the human body by radioactive scans.

Python – Python is a high-level, object-oriented, open source, general purpose programming language.

Ubuntu – Ubuntu is a open source Linux distribution based computer operating system.

## ABSTRACT

Autism Spectrum Disorder (ASD) is part of the developmental disabilities. There are numerous symptoms for ASD patients, including lack of abilities in social interaction, communication obstacle and repeatable behaviors (Centers for Disease Control and Prevention., 2020c). Meanwhile, the rate of ASD prevalence has kept rising by the past 20 years from 1 out of 150 in 2000 to 1 out of 54 in 2016 (Centers for Disease Control and Prevention., 2020a). In addition, the ASD population is quite large. Specifically, 3.5 million Americans live with ASD in the year of 2014, which will cost U.S. citizens \$236-\$262 billion dollars annually for autism services (Buescher, Cidav, Knapp, & Mandell, 2014). So, it is critical to make an accurate diagnosis for preschool age children with ASD, in order to give them a better life. Instead of using traditional ASD behavioral tests, such as ADI-R, ADOS, and DSM-IV, we applied brain MRI images as input to make diagnosis. We revised 3D-ResNet structure to fit 110 preschool children's brain MRI data, along with Convolution 3D and VGG model. The prediction accuracy with raw data is 65.22%. The accuracy is significantly improved to 82.61% by removing the noise around the brain. We also showed the speed of ML prediction is 308 times faster than behavior tests.

## CHAPTER 1. INTRODUCTION

In this chapter, we provided a big-picture of the study. We introduced some background information about autism spectrum disorder (ASD) and machine learning (ML). In addition, we proposed the disadvantages of behavior tests and addressed the gap of applying machine learning method to ASD diagnosis. Then, we concluded two main research questions and listed the assumptions, limitations and delimitation of the study. Finally, there is a summary section at the end of the chapter.

### 1.1 Background

The origin of the word "autism" is taken from a Greek word "autos", which means "self". It describes the disability of social, communication and behavioral interaction, which makes the ASD patients become "isolated self". ASD is one of the neurological developmental disorders with several symptoms, such as not to look at or listen to people, having disability of conversation with back and forth, repeat certain behaviors, not wanting to change in their daily activities and having overly focused interests. In fact, the prevalence of ASD has kept increasing since the 1970s. The study proposed by Maenner Mj, Shaw KA, Baio J, et al. (2020) showed the ASD prevalence in the year of 1970s is 1 in 10000 people, but the prevalence increased to 1 in 54 people in the year of 2016, which means the ASD prevalence increased 169 times in 50 years (Maenner MJ, 2020). Moreover, the population of ASD patients is huge, with approximately 3.5 million Americans living with ASD. The U.S government has to pay \$236-\$262 billion dollars annually for autism services (Buescher et al., 2014). The good news is there are some treatment solutions for ASD patients (Arnold et al., 2012). The earlier ASD patients get treated, the better chance they are in successful treatment. Thus, a early age diagnosis with high accuracy is needed to benefit ASD community.

Machine Learning (ML) is an interdisciplinary subject of computer science and statistics, which applied different models to make predictions. It is a subset of artificial intelligence. Machine Learning method becomes ever more popular due to the improvement of computer hardware and the explosion of data size. Machine learning already showed its power for some areas of clinical practice, such as identifying cardiovascular abnormalities (Martin-Isla et al., 2020), detecting musculoskeletal injuries (Boissoneault, Sevel, Letzen, Robinson, & Staud, 2017), diagnosing neurological diseases (Orrù, Pettersson-Yeo, Marquand, Sartori, & Mechelli, 2012), aiming for thoracic illness (Retson, Besser, Sall, Golden, & Hsiao, 2019) and locating for cancers (Kourou, Exarchos, Exarchos, Karamouzis, & Fotiadis, 2015). Also, there are lots of studies already showing that ML could be an effective method to solve image-based clinical challenges. The image-based clinical problem means medical staff and researchers have to use medical imaging techniques in order to make a diagnosis. Some different medical imaging techniques includes X-ray, MRI, nuclear medicine, ultrasound etc. In this study, we focused on MRI, which is the most popular and common technique to display brain structures, which is the main cause of neurological diseases. There are lots of studies showing excited results by applying ML to MRI imaging to make a diagnosis, which will be addressed in the literature review section.

## 1.2 Problem Statement

The traditional ways to diagnose ASD are by using behavioral tests, such as ADI-R, ADOS, and DSM-IV. Even though the ASD diagnosis could be made by experienced professionals at the age of 2, the reality is that most children receive their official diagnosis after few years (Lord et al., 2006). Also, most screening tools take 2-15 minutes to accomplish, which is a challenge for preschool children (Committee on Children with Disabilities and others, 2001; Glascoe, 1998). Thus, it is necessary to apply ML methods for ASD diagnosis due to the following reasons:

- ASD behavioral tests only have approximately 70% of accuracy for diagnosis for preschoolers, and the accuracy is even lower for younger children (Committee on Children with Disabilities and others, 2001; Glascoe, 1998). ML approaches could be self-evolved for higher accuracy while data size is increasing. A ML approach could make better predictions in the early stage than behavioral tests.
- ASD behavioral tests do not have capacity to point out the pathology of ASD and find biomarkers, but ML approach has the potential to find pathology for ASD.
- ASD behavioral tests need to be operated by experienced professionals, while ML approach does not need professionals to involve.
- ASD behavioral tests require 2-15 minutes to be accomplished and most of final diagnosis are made within 3-4 years, but ML approach only needs a few seconds to make prediction and 1-2 hours to gather MRI images.

Therefore, it is useful to apply ML methods on ASD diagnosis, in terms of accuracy and efficiency. There are several studies showing differences of grey matter thickness between ASD brain and normal brain, which ensures the high possibility that CNN-based models could make good predictions (Hutsler, Love, & Zhang, 2007). In addition, Hazlett et al. (2017) proposed that deep learning could be useful to predict ASD in infants. So, this study aims to implement a CNN-based network for preschoolers ASD diagnosis by reading brain MRI images with high accuracy and speed.

### 1.3 Scope of the Problem

The scope of data selection is that ASD is a developmental disorder, so the patient's age could vary from new born babies to very old people. However, a study proposed by Giedd described that the total size of a six years old child's brain could reach approximately 90% of its adult size (Giedd, 2004). we have been given access to anonymized 110 preschooler from three years old to six years old as research objects, and ignored the rest of the ages. In addition, we only focused on general ASD classification, which means we did not classify different types of ASD, such as high functional autism (HFA), low functional autism (LFA) and pervasive developmental disorders (PDD).

The scope of machine learning method is that we mainly focused on Convolutional Neural Network (CNN) structure, not only because CNN is the most accurate and efficient machine learning model for image-based problems, but also because CNN structure showed some lovely results for other neurological diseases prediction, such as Alzheimer's disease (Bernal et al., 2019; Lin et al., 2018). In addition, we only used each individual's 156 MRI raw data as input for the machine learning model. We utilized other medical visualization tools to complete the 3D reconstruction tasks. So, this study will not discuss MRI image segmentation and 3D reconstruction.

## 1.4 Significance of the Problem

With the development of medical science, most physical diseases can be examined and treated properly. On the contrary, there are lots of mental diseases that can not be identified and prevented by modern medicine. Markram stated that ASD is a neurological disorder that is affected by gene and hereditary. The probability of identical ASD occurred in monozygotic twins is approximately 60%, nevertheless siblings only have 2% - 7% of risk (Markram, Rinaldi, & Markram, 2007). Besides, the ASD prevalence is increasing dramatically over the past decade (Volkmar, Lord, Bailey, Schultz, & Klin, 2004). Thus, ASD has heredity and it will discourage patients' normal life. Patients with ASD have symptoms, including lack of sociability, communication disabilities, shortage of imagination and repetitive behaviors (DSM-IV, 1994; Markram et al., 2007). So, ASD is a severe mental disease, which can affect an individual's habitual life.

On the other hand, ASD can be intervened on at an early stage, which means that it could be treated if doctors can make a diagnosis at an early stage. In addition, ASD patients' brain grey matters are approximately 2.5 percent thicker than normal people, but this difference will decrease while age is increasing. In other words, ASD detection is easier to be detected in adolescents rather than adult (Giedd, 2004; Hutsler et al., 2007). All these studies show the significance of detecting ASD at an early age. Besides, there are some treatments that can improve ASD patient's social ability, such as sulforaphane and melatonin (Garstang & Wallis, 2006; Singh et al., 2014). Therefore, ASD can be processed if it has been detected correctly and promptly.

It is also important to apply machine learning methods to ASD diagnosis, not only because behavioral tests are complicated and expensive, but also because it has potential to find biomarkers of ASD pathology. Even though there are several ASD behavioral tests screening tools, they have a few significant limitations. For example, screening tools only have 70% accuracy for preschoolers ASD diagnosis (Committee on Children with Disabilities and others, 2001; Glascoe, 1998). Another limitation of behavioral tests is that they usually take 2 to 15 minutes to be accomplished by professionals (Dobrez et al., 2001; Glascoe, 1998). On the contrary, machine learning methods could make the diagnosis in a few seconds with higher accuracy.

### 1.5 Statement of Purpose

This research examines if CNN-based ML models could be applied to brain MRI images for preschoolers ASD diagnosis with higher accuracy and speed. This study will utilize this approach and compare the diagnostic accuracy and diagnose speed with traditional ASD behavioral testing.

### 1.6 Research Questions

This study investigated the possibility of applying machine learning methods to brain MRI images for preschoolers ASD detection, especially by CNN-based structure. This study also discussed how to implement a binary classifier to predict a new patient with ASD in high accuracy. In other words, this study aims to answer two major questions:

1. Could the use of CNN-based structure detect preschoolers with ASD by brain MRI images?
2. Could the proposed CNN-based structure has higher accuracy and speed than behavioral tests?

### 1.7 Assumptions

The assumptions of this study are:

- We assumed that all data is valid, which means all individuals who labeled as ASD are ASD patients. And all individuals who are labeled as TD are not ASD patients.
- We assumed that there are differences between a healthy brain and illness brain with ASD, in terms of the thickness and volume of grey matter.
- We assumed each layer of MRI scan does affect adjacent layers for ASD detection, which means we need to consider temporal information for the study.

## 1.8 Limitations

Based on the design of the study, three limitations are shown as below:

- The MRI data set only contains 110 individuals. With the 8:2 training-testing data split ratio, there are only 22 individuals in the testing set, which is relatively small.
- We do not have a solid medical background.
- The most literature reviews and statistical data that used in this research are based on the United States, but the dataset is collected in China. Although the pathology of ASD could be the same between U.S. and China, the result may misinterpret for the entire ASD population.

## 1.9 Delimitation

The delimitation of this study includes:

- The version of the operating system is Linux Ubuntu 18.04.
- The hardware configuration is Intel i9-9900k (3.6GHz), Nvidia Titan RTX GPU (24GB GDDR6), 64GB RAM.
- The version of Freesufer is V6.0.0.
- The version of Freeview is 2.0.
- The dataset used in this study is a private dataset, which is collected by a group of professional radiologists in a top ranking hospital in Beijing, China.

## 1.10 Summary

This chapter is mainly focused on the big picture of this study. The "background" part introduced some key information about ASD and ML. The "problem statement" part discussed the limitations of current ASD diagnosis methods. The "scope of the problem part" introduced both data selection scope and machine learning method scope. The "significance of the problem" part discussed the importance of applying ML, which could improve the current ASD diagnosis performance. The "statement of purpose" part indicated the goal of this research. The "research question" part summarized two main research questions. The following "assumption", "limitation" and "delimitation" parts introduced our assumptions, as well as some limitations and the scope of software, hardware and dataset.

## CHAPTER 2. REVIEW OF LITERATURE

In this chapter, we introduced some important studies of ASD, such as the ASD incentive, the differences of Grey Matter between ASD brain and normal brain and the major ASD diagnosis. In addition, we discussed some Machine Learning models that could be applied to this study, such as Convolutional Neural Network (CNN), Visual Geometry Group (VGG) and Residual Network (ResNet). This chapter also included a section of contemporary studies of applying ML to ASD diagnosis.

### 2.1 Overview of Autism Spectrum Disorder

ASD is a type of neural disease and the term of "Autism" was introduced by Leo Kanner in 1943. Markram et al. (2007) proposed that ASD can be recognized before three years old and get worse increasingly in the rest of life. The prevalence of ASD has been growing since the 1970s, and the overall population of ASD in the U.S. is around 3.5 million (Buescher et al., 2014; Maenner MJ, 2020). ASD patients have two main types of symptoms, which are lack of social communication/interaction abilities, and restrictive/repetitive behaviors. Some most common symptoms of social communication and interaction disability are:

- Do not talk and respond to other people.
- Do not have facial expressions and eye contacts.
- Avoid any types of physical interaction with other people, like shaking hands or kissing.
- Can not understand other people's feelings. ASD patients always lack empathy. (Centers for Disease Control and Prevention., 2019)

Some symptoms of restrictive and repetitive behaviors are:

- Organized personal tasks very well and repeated every time, felt sad if there were any changes.
- Very focus on one interest and do not want to change.

- Repeated some words and sentences. (Centers for Disease Control and Prevention., 2019)

Another interesting fact is that ASD patients are not always having low Intelligence Quotient (IQ). Actually, some ASD patients have much higher IQ than normal people. ASD patients can be separated by two groups, which are high functional autism (HFA) and low functional autism (LFA). HFA described the ASD patients who obtained a nonverbal IQ of 70 or above and a verbal IQ of 70 or above; LFA described the ASD patients who obtained a nonverbal IQ of 69 or below and a verbal IQ of 69 or below. Besides, another neural disease has very similar symptoms as ASD, which is called pervasive developmental disorders (PDD). Markram et al. (2007) proposed an alternative hypothesis for ASD, and they believed that ASD patients have hyper-functionality. They indicated that ASD patients are too sensitive to gather more information from other people, which makes ASD patients only live in their own world. However, we did not classify HFA, LFA and PDD into different groups in this study.

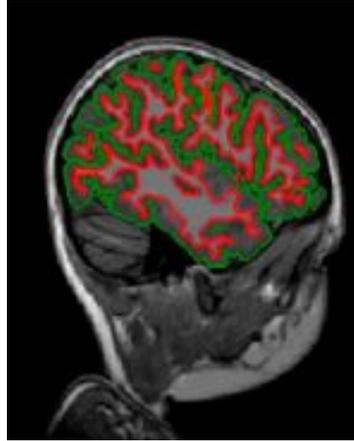
### 2.1.1 ASD Incentive

Although medical professionals are still trying to find exact incentives for ASD, there are studies showing some potential incentives for ASD. The first approach of ASD incentive belongs to genes. Hallmayer et al. (2011) proposed the consistency of having ASD for monozygotic male twins is 0.77, as well as 0.5 for monozygotic female twins. Also, the gender of newborns can affect ASD prevalence. Loomes, Hull, and Mandy (2017) proposed that the ASD prevalence in male is three times more than female. In addition, some studies showed that low birth weight may be associated with ASD prevalence (Itzhak, Lahat, & Zachor, 2011; Mann, McDermott, Bao, Hardin, & Gregg, 2010).

Another approach of ASD incentive is the mutation of embryogenesis, which means gravidas have toxic exposure during pregnancy, such as misoprostol (Bandim, Ventura, Miller, Almeida, & Costa, 2003), alcohol (Nanson, 1992), VPA (Moore et al., 2000; Rasalam et al., 2005), and thalidomide (Strömmland, Nordin, Miller, Akerström, & Gillberg, 1994).

### 2.1.2 The Differences of Grey Matter

Grey Matter is one of the cerebrum portions of brain, which is located between pial and white matter. The most of brain neuronal cells is located at Grey Matter, which endures the ability of Grey Matter to dominate people's memories, emotion, muscle controlling etc Miller, Alston, and Corsellis (1980). See Figure 2.1



*Figure 2.1.* Grey Matter is located between the green line and red line. The green line represents the pial-grey matter boundary, and the red line represents grey matter-white matter boundary.

Some studies showed ASD patients have thicker Grey Matters in some specific areas. Hutsler et al. (2007) proposed ASD patients have 2.5% thicker cortical area than normal people. They also stated that there are 2% differences in Frontal and Parietal, as well as 2.5% differences in Temporal area between ASD and TD group. Chen, Jiao, and Herskovits (2011) proposed ASD and TD group has different grey matter volumes. They were using voxel-based morphometry to determine evidence for increased grey matter volumes and surface-based morphometry to find changes in the Frontal, Temporal, and Parietal area in ASD group. It is always better to use the thickness and volume of Grey Matter to explore ASD in the early age, because the Grey Matter thickness differences between ASD and TD group gets smaller when ages increased (Hutsler et al., 2007). Giedd (2004) proposed that the brain size of six years old children can get 90% of its matured size. So we used 3 years old to 6 years old children brain MRI scans as our dataset.

### 2.1.3 ASD Diagnosis

The most traditional and reliable way to diagnose ASD is fulfilling ADI-R, ADOS, or DSM-IV criteria. According to DSM-IV criteria, a typical ASD patient has the following symptoms:

- Hard to communicate and interact with other people.
- Repeated behaviors and limited interests.
- Discouraged social behaviors, such as disability in school, work etc. DSM-IV (1994)

Lord et al. (2006) proposed that children with ASD can be diagnosed at the age of 2 by a professional. However, most children with ASD can not get their final diagnosis after many years. In addition, the procedure of behavioral tests is very complicated. Parents also have to be involved in the test, because children are too young to understand the question. In this situation, parents may fulfill the wrong criteria, depending on how well they observe their children and how well they understand the question. Also, behavioral tests can take up to 15 minutes to be completed, which is much longer than Machine Learning prediction. The comparison of behavioral tests (Figure 2.2) and Machine Learning prediction (Figure 2.3) workflow is shown below:

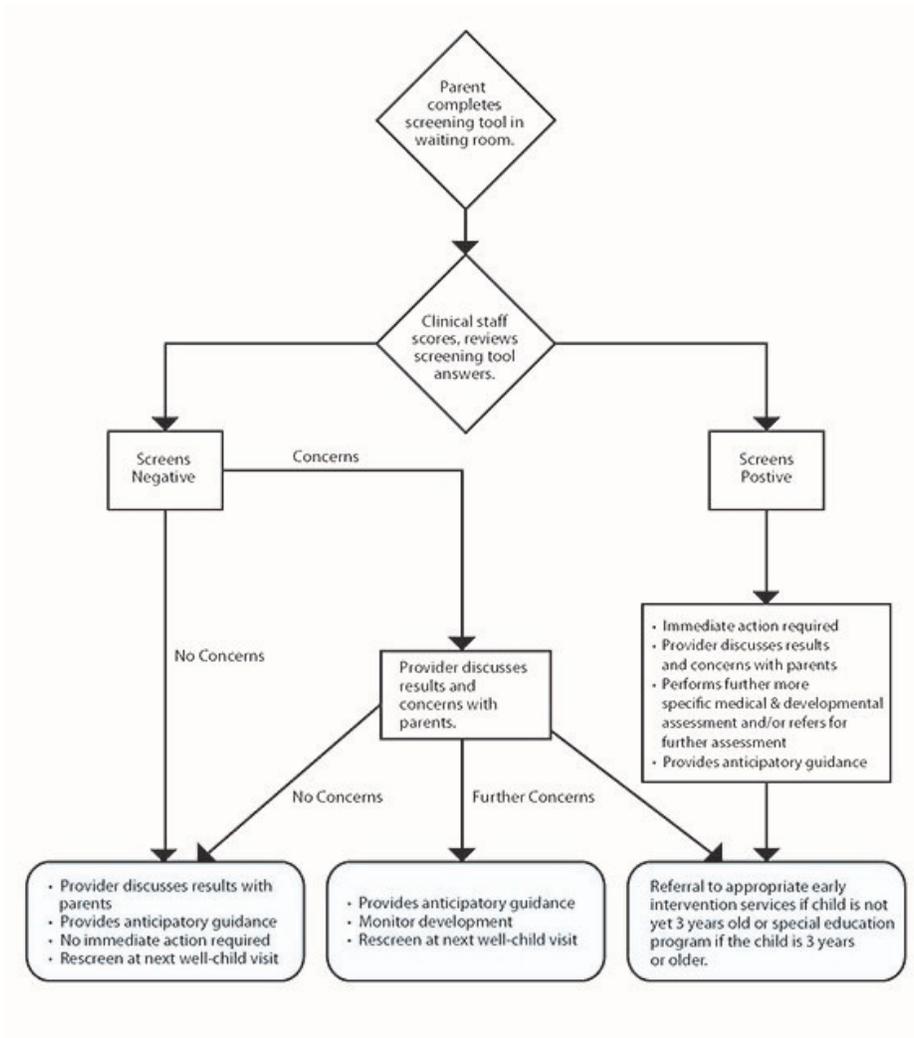


Figure 2.2. The workflow of DSM-IV for ASD diagnosis (Centers for Disease Control and Prevention., 2020b).

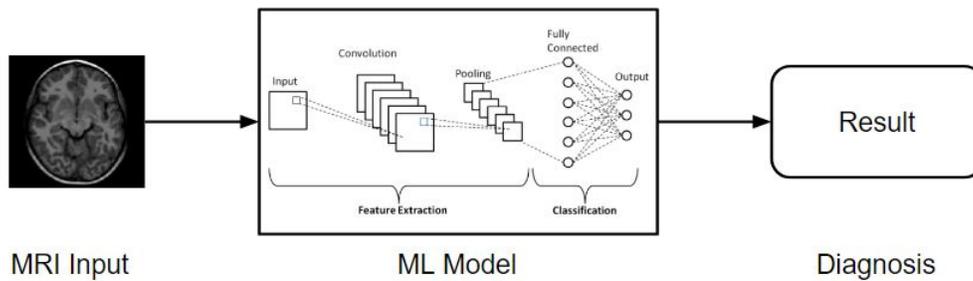


Figure 2.3. The workflow of ML method for ASD diagnosis.

## 2.2 Overview of Different Machine Learning Models

In 1959, the terminology of "Machine Learning" was proposed by Arthur Samuel (Samuel, 1959). Machine Learning (ML) is an interdisciplinary field of scientific study, combined with computer science and statistics. ML applies different mathematical algorithms and statistical models to different tasks by using computers. The purpose of ML is to make predictions based on sample data, which is also called "training data". As computer and internet technology develops, the data grows exponentially in the past 20 years, which promoted the evolution of Machine Learning. Therefore, more and more fields of study are starting to use Machine Learning methods, in order to generate predictions faster and more accurately.

Artificial Neural Network (ANN) is one of the ML model, inspired by biological neural network structure in animals. The prototype of ANN can be traced back to 1940s. The first ANN with multiple layers is proposed by Ivakhnenko and Lapa in 1965 (Ivakhnenko & Lapa, 1966, 1967; Schmidhuber, 2015). An ANN is constructed by multiple connected nodes, which is called neurons. These neurons could receive signals from the previous layer and process the information, as well as send the result signal to the next layers. The fundamental functionality of ANN approach is very similar to the functionality of the human brain. The original purpose to design an ANN is letting computers solve some questions like humans, but it turns out that ANN is an efficient tool for several tasks, such as image processing, speech recognition, gaming AI etc. In this study, we focused on Convolutional Neural Network structures, because the data input for this research is MRI images, and CNN shows an extraordinary performance on image processing.

### 2.2.1 Convolutional Neural Network Overview

Convolutional Neural Network (CNN) is a category of ANN, which contains at least one convolutional layer. The CNN structure is commonly applied to computer vision areas. There are several models built on top of CNN structure. In 1989, LeCun et al. (1989) presented an ANN model called LeNet, which is the first CNN model that applied on computer vision. CNN has a shift invariant property, letting CNN makes the correct classification even if the object is shifted (Zhang et al., 1988). CNN's approach is famous for computer vision, so it wins the first place multiple times on one of the most famous computer vision competitions – ImageNet Challenge (ILSVR). In other hands, ILSVR encouraged the evolution of CNN, so lots of reliable and outstanding CNN models are published in ILSVR. See Table 2.1.

Table 2.1. *Some famous models in ILSVR*

Year	Network/Model
2012	AlexNet (1st Place)
2014	GoogleNet (1st Place)
2014	VGG (2nd Place)
2015	ResNet (1st Place)

Also, the model has been becoming much more complex as time goes on, which has more parameters and more layers. The reason for this phenomenon is that deeper CNN structures could extract higher dimensional features, which could increase the classification accuracy. In ILSVR, AlexNet was used in 2010 with only 8 layers, as well as ResNet was proposed in 2015 with 152 layers. See Figure 2.4.

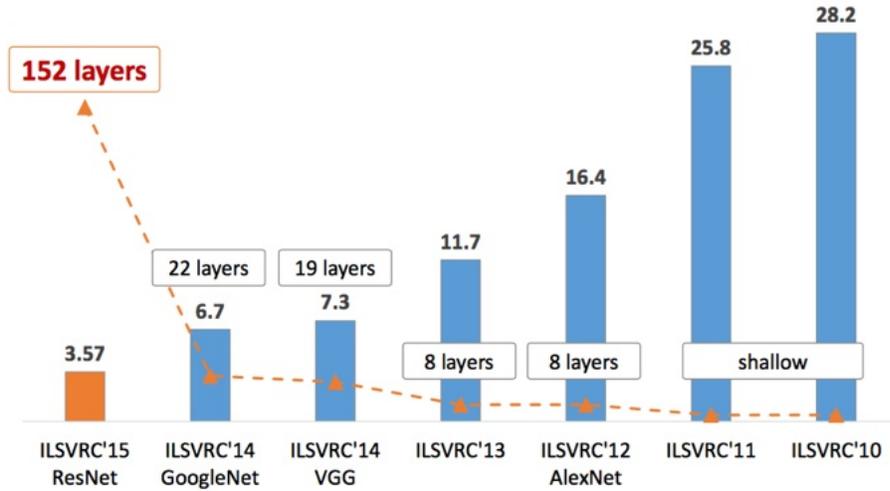


Figure 2.4. Different number of layers used in different models.

### 2.2.1.1 CNN Structure

CNN model is constructed by three main components, which are convolution, pooling and fully connected layer. Convolution layer can extract features from original images by convolution kernels. Pooling layer introduces the invariance properties and dimension reduction of CNN. Fully-connect layer can summarize all features representation to one value by applying a activation function. CNN structure is shown as Figure 2.5.

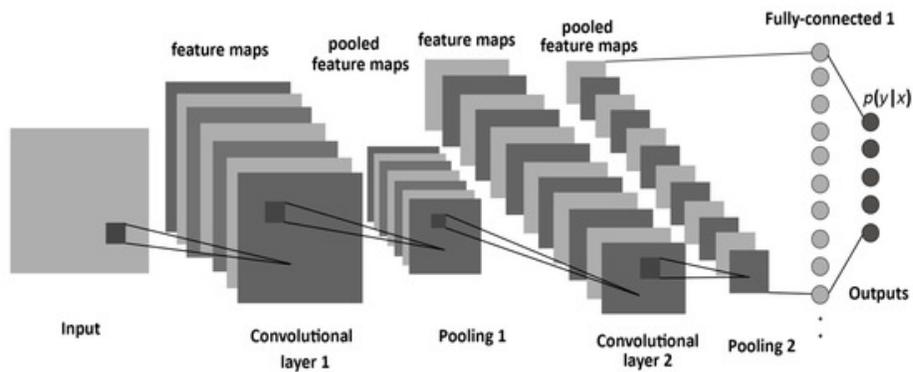


Figure 2.5. Convolutional Neural Network structure (Albelwi & Mahmood, 2017).

The input layer in CNN could take a tensor as input with the size of:

$$Tensor\ Size = (number\ of\ image) * (image\ width) * (image\ height) * (image\ depth) \quad (2.1)$$

In convolution layer, a convolution kernel is used to generate the feature map by the following formula:

$$\int_{-\infty}^{\infty} f(\tau)g(x - \tau)d\tau \quad (2.2)$$

In formula (2.2),  $f(x)$  is source pixels, and the summation of  $f(x)$  is original input images. The summation of  $g(x)$  represents the convolution kernel. The convolution kernel has a size of  $n*n$ . In order to generate the feature map, each source pixel should perform a dot multiply with corresponding pixel on convolution kernel and make a summation. The result was recorded as the destination pixel on the feature map.

Next, feature map goes to a connected pooling layer, which is also called a subsampling layer. Pooling layer can select the most significant feature and ignore noise features. The pooling process could reduce the possibility of overfitting, as well as the number of parameters. In addition, the pooling process also enables the property of shift invariant, which could improve the robustness of the model (Zhang et al., 1988). There are two main pooling methods, which are Max Pooling and Average Pooling (Ciregan, Meier, & Schmidhuber, 2012; Mittal, 2018). It is possible to add as many convolution layers and pooling layers as possible, unless the dimension of the pooled feature map becomes [1x1].

The last step of CNN is to classify all samples to different classes by using fully connected layer. The activation could be calculated with matrix multiplication, and the bias was added in this step. The fully-connected layer is beneficial to prevent object shifting, because it integrates all feature representation into one value.

### 2.2.1.2 Conv3D: 3 dimensional CNN

3D Convolutional Network (Conv3D) was proposed in 2012 in order to fulfill the human action recognition tasks (Ji, Xu, Yang, & Yu, 2012). The advantage of Conv3D approach is the ability to learn spatiotemporal features, which can not be accomplished by 2D CNN (Tran, Bourdev, Fergus, Torresani, & Paluri, 2015). Thus, Conv3D is commonly used in video-based computer vision tasks, such as irregularities detection (Boiman & Irani, 2007), action detection and recognition (Laptev, 2005; Over et al., 2013), medical image segmentation (Jain et al., 2010; Turaga et al., 2010) etc. The main differences between 2D CNN and Conv3D is convolutional layer and pooling layer in terms of structural design. Conv3D has better ability and performance to learn temporal information than 2D CNN. In simple words, 2D CNN could only output an image after convolutional layer and pooling layer, but Conv3D could output a volume with spatio-temporal information (Tran et al., 2015). Although applying 2D CNN on multiple frames is possible, it only generates a single image output for each frame. See Figure 2.6.

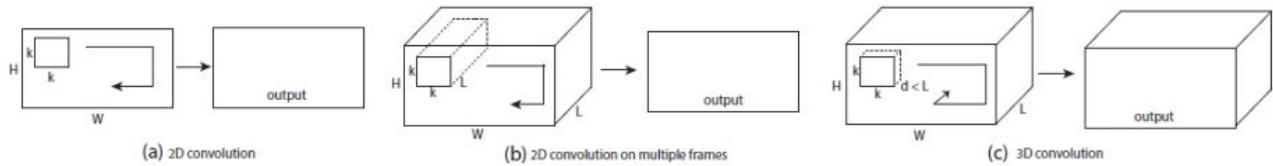


Figure 2.6. (a)2D convolution with kernel size  $k*k$  (b)2D convolution on L-frames with kernel size  $k*k$  (c)3D convolution on L-frame with kernel size  $k*k*d$

In this study, the data input is a set of objects with 157 MRI images for each object. For each object, the 157 MRI images were collected sequentially with 1mm interval. Because each brain region could across several MRI images, so the ASD detection result has to rely on multiple adjacent MRI scans comprehensively. Therefore, the temporal information for this study is very crucial.

### 2.2.2 Visual Geometry Group

Visual Geometry Group (VGG) was proposed by Karen and Zisserman in 2014 (Simonyan & Zisserman, 2014). The most famous VGG networks are VGG-16 and VGG-19, which won the second place in ILSVR 2014. VGG-16 and VGG-19 do not have differences in terms of structural design, but VGG-19 has more hidden layers than VGG-16. One of the main innovation of VGG network is that VGG used two 3 by 3 kernel to replace a 5 by 5 kernel, and used three 3 by 3 kernel to replace a 7 by 7 kernel in AlexNet (Krizhevsky, Sutskever, & Hinton, 2012). See Figure 2.7.

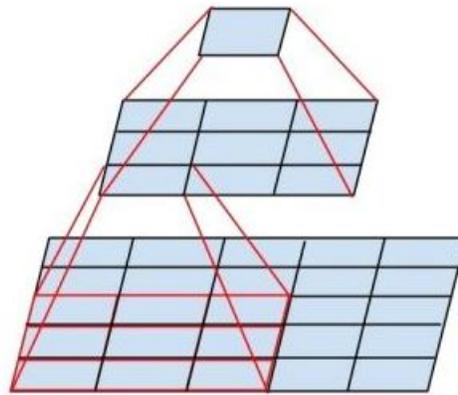


Figure 2.7. Demonstration of how to use two 3\*3 kernel to replace one 5\*5 kernel.

This design not only reduced the number of parameters, but also increased the depth of the neural network effectively. In terms of the number of parameters, one 5 by 5 kernel can generate 25 parameters, but two 3 by 3 kernel only generate  $2*3*3=18$  parameters, which saves computational power with efficiency. Similarly, one 7 by 7 kernel can generate 49 parameters, but three 3 by 3 kernel only generate  $3*3*3=27$  parameters. The structure of VGG is very simple, so it is very easy to be implemented. All convolution layers in VGG use kernel size of 3 by 3, and all pooling layers use max-pooling with the size of 2 by 2. See Figure 2.8.

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input ( $224 \times 224$ RGB image)					
conv3-64	conv3-64 <b>LRN</b>	conv3-64 <b>conv3-64</b>	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 <b>conv3-128</b>	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 <b>conv1-256</b>	conv3-256 conv3-256 <b>conv3-256</b>	conv3-256 conv3-256 conv3-256 <b>conv3-256</b>
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 <b>conv1-512</b>	conv3-512 conv3-512 <b>conv3-512</b>	conv3-512 conv3-512 conv3-512 <b>conv3-512</b>
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 <b>conv1-512</b>	conv3-512 conv3-512 <b>conv3-512</b>	conv3-512 conv3-512 conv3-512 <b>conv3-512</b>
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Figure 2.8. VGG network structure. Column D represents the structure of VGG-16. Column E represents the structure of VGG-19.

Another contribution of VGG is that the author approved increasing the depth of neural networks could improve the performance, which inspired other researchers to design very deep CNN.

However, the main limitation of VGG is that it generates too many parameters in fully connected layer. There are more than four thousands of parameters in both the first and the second fully connected layer. The third fully connected layer contains one thousand of parameters. This limitation causes more memory consumption and longer training time. Another limitation of VGG is that if the network is too deep, the performance starts to decrease, which is called a degradation problem. Because it is very difficult to optimize deeper layers and avoid vanishing gradients and exploding gradients problems.

### 2.2.3 Residual Networks

Residual Networks (ResNet) was proposed by He et al. in 2015, and it won the first place in ILSVR 2015. The main purpose of designing ResNet is to solve vanishing gradients problem and degradation problem. Therefore, He et al. (2016) proposed a new structure, called residual block. See Figure 2.9

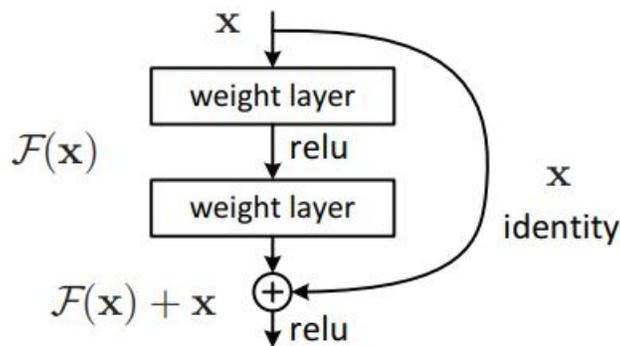


Figure 2.9. A building block structure from ResNet (He et al., 2016).

In the building block,  $x$  is the input, and  $F(x)$  represents the residual function. There is an identity shortcut connection, which could jump one or multiple layers. The features can be learned by building block could be denoted as

$$H(x) = F(x) + x \tag{2.3}$$

The advantage of this design is that deeper layers at least have the same performance  $x$  as previous layers, if residual function  $F(x)$  does not learn any new features or degrade performance. The building block could be represented by the formula (He et al., 2016):

$$y = F(x, \{W_i\}) + x \quad (2.4)$$

In this equation,  $x$  and  $y$  are inputs and outputs feature maps for the current residual block. The function  $F(x, \{W_i\})$  is the residual mapping to be learned by ResNet. In addition, He et al. (2016) also proposed another type of residual block called “bottleneck” building block, which is commonly used in deeper ResNet. The difference between a regular building block and “bottleneck” build block is that a regular building block only has two convolutional kernels with size of 3 by 3 and 64 channels, but “bottleneck” building block has three convolutional kernels. The first convolutional kernel for “bottleneck” block has size of 1 by 1 with 64 channels, as well as the second kernel has the size of 3 by 3 with 64 channels and the third kernel has the size of 1 by 1 with 256 channels. Regular build blocks are commonly used in “shallow” ResNet, such as ResNet-34, and “bottleneck” blocks are often used in deeper networks, such as ResNet-50/101/152. The Figure 2.10 represents both regular build block and “bottleneck” block structure.

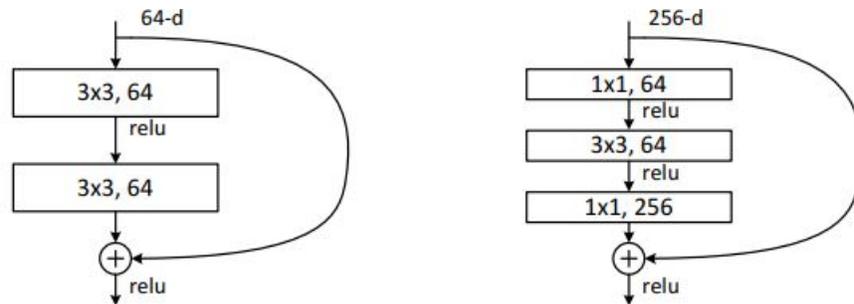


Figure 2.10. Left: the regular building block; Right: the “bottleneck” block (He et al., 2016).

The “shortcut” feature ensures ResNet has much more layers than VGG-19. Besides, this feature is also applied by other networks, such as ResNeXt (Xie, Girshick, Dollár, Tu, & He, 2017) and DenseNet (Huang, Liu, Van Der Maaten, & Weinberger, 2017).

### 2.3 Current Study of Applying ML to ASD Diagnosis

There are lots of studies that showed that researchers started to apply Machine Learning methods to ASD diagnosis in different ways. Some studies take behavioral tests and apply statistical ML models to prediction, while other studies use brain MRI images as input and apply image-based ML models to make prediction. For the second approach, the Machine Learning method aims to solve two typical types of problems in clinical practices based on MRI, which are magnitude images segmentation and classification (Lundervold & Lundervold, 2019). Heinsfeld, Franco, Craddock, Buchweitz, and Meneguzzi (2018) proposed their Deep Neural Network (DNN) approach for ASD detection, and ended with 70% accuracy. However, their DNN approach takes around 33 hours, which is a lot longer than regular behavioral tests. Besides, 70% accuracy does not show a big improvement compared with behavioral tests. Li, Dvornek, Zhuang, Ventola, and Duncan (2018) proposed a two-stage architecture with DNN to find biomarkers for ASD. They used two CNN in parallel with different inputs. They fed raw fMRI data into the first well trained CNN model and fed fMRI image with corrupted Region of Interest (ROI) into the second well-trained CNN model. Researchers used distribution analysis to interpret which ROI is the biomarker for ASD. The two-stage approach achieved 87.1% accuracy on their private dataset (Di Martino et al., 2014; Smith et al., 2004). However, the training-testing split ratio is not common, which is 92:8, so they only have 10 subjects in the testing set. Hence, each testing subject can affect the final accuracy by 10%. In other words, their accuracy could increase to approximately 100% if their model made one more correct prediction, as well as decrease to around 70% if their model made one more incorrect prediction. In addition, they did not mention how they corrupt ROI on fMRI dataset, which reduces the reliability of their result. Some other studies also showed some excited results. Hazlett et al. (2017) applied a DNN to predict infants with ASD at high risk, and they got 94% prediction accuracy. Liu et al. (2015) proposed a k-means clustering model to predict ASD by patients eye movement with 80.33% accuracy in the adult group with motion and fusion detection and 86.89% accuracy in the child group with fusion detection. These studies can be summarized in Table 2.2.

Table 2.2. *Different studies of applying ML to ASD prediction.*

Research	Structure	Accuracy(dataset)	Biomarkers
Heinsfeld et al. (2018).	DNN	70% (ABIDE)	No
Li et al. (2018).	2CC3D	87.1% (private)	Yes
Hazlett et al. (2017).	DNN	94% (private)	Yes
Liu et al. (2015).	SVM	86.89% (private)	No

After reviewed these studies, we found some homogeneous limitations:

- Most of researches used private dataset that other researchers are not able to access. The quality of public dataset, such as ABIDE, is lower than private dataset, in terms of patients' age range, male-female ratio and ASD-TD ratio.
- Only a few researchers introduced the design and implementation of ML models.
- Almost all studies used a relatively small dataset as input, which made the size of the testing set very small.

#### 2.4 Summary

In this chapter, we first reviewed several studies about Autism Spectrum Disorder and discussed the definition, symptoms, incentives and diagnosis of ASD. Then, we talked about some Machine Learning models, including CNN, VGG and ResNet. We also reviewed some research which applied Machine Learning methods to ASD diagnosis.

## CHAPTER 3. METHODOLOGY

In this chapter, we proposed the hypothesis and experiment environment for the study. In addition, we described our dataset, Machine Learning models and algorithms. We discussed how to evaluate and analyze our experiment result in the end.

### 3.1 Hypothesis

The null hypothesis and alternative hypotheses for this study are shown below:

$H_0$ :

- The use of ML method does not improve the accuracy of ASD diagnosis by using MRI images.
- The use of ML method does not increase the speed of ASD diagnosis by using MRI images.

$H_a$ :

- The use of ML method does improve the accuracy of ASD diagnosis by using MRI images.
- The use of ML method does increase the speed of ASD diagnosis by using MRI images.

### 3.2 Experiment Environment

This section introduces the hardware setups and software configurations for the study.

### 3.2.1 Hardware

It is very time consuming for the Machine Learning model training process in general speaking. Especially, the more layers a model has, the more parameters will be generated, which can consume lots of computational powers. A good thing is that GPUs started to add more machine learning features to help data scientists overcome this bottleneck. In this study, we used two hardware setups parallel in order to speed up the experiment progress. To be clarified, we only split the original dataset and run each dataset individually by one hardware setup, but we did not use parallel computing techniques. See Table 3.1 and Table 3.2.

Table 3.1. *The first hardware setup.*

Components	Configuration
CPU	Intel Core i9-9900k @3.60GHz
RAM	DDR4 64GB
GPU	Nvidia Titan RTX, GDDR6 24GB

Table 3.2. *The second hardware setup.*

Components	Configuration
CPU	Intel Core i9-9900k @3.60GHz
RAM	DDR4 64GB
GPU	2x Nvidia GeForce RTX 2080Ti, GDDR6 11GB

### 3.2.2 Software

Ubuntu 18.04 LTS was installed for both computers. In addition, Freesufer V6.0.0 and Freeview 2.0 are installed on both computers for MRI 3D reconstruction and segmentation purposes. All Machine Learning models are implemented and tested by Python 3.7. See Table 3.3.

Table 3.3. *The software specification.*

Software	Version
OS	Ubuntu 18.04 LTS
Freesufer	V6.0.0
Freeview	2.0
Python	3.7

Several Machine Learning packages were imported. See Table 3.4.

Table 3.4. *Imported ML packages.*

Packages	Version
PyTorch	1.1.0

### 3.3 Data

#### 3.3.1 Data Source and Anonymization

The data was gathered from GE 750 3T MR machine by a group of professional radiologists in Beijing, China. The data format is Digital Imaging and Communications in Medicine (DICOM). The collecting procedure was in conformity with the Chinese law. The data does not include any sensitive information, including name, date of birth, official IDs etc. The dataset came with only an object ID number and a label, indicating whether objects are ASD or TD.

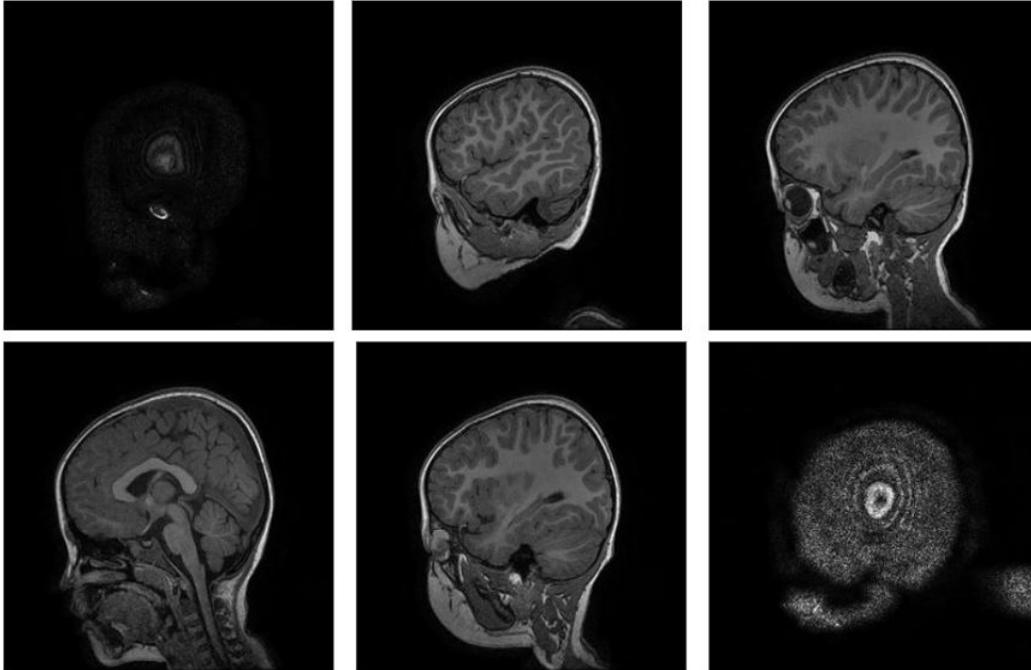
### 3.3.2 Data Description

We used T1-weighted preschooler’s brain MRI images as the input for Machine Learning models. The dataset has 110 objects who are preschoolers from 3 years old to 6 years old. Each object was labeled as autism spectrum disorder (ASD) or typical development (TD). There are 61 objects in the ASD group, as well as 49 in the TD group. In addition, the male to female ratio of ASD group is 2:1, and 11:10 in the TD group, because the ASD prevalence of male is higher than female (Loomes et al., 2017). Each object in the dataset has exactly 156 slices of T1-weighted brain MRI scans. See Table 3.5.

Table 3.5. *The statistics of dataset.*

	ASD Group	TD Group
Objects	61	49
Ave. Age (years old)	4.56 +/- 0.97	5.13 +/- 0.82
Male-Female Ratio	2:1	11:10

The original data type of MRI images is DICOM format. But Machine Learning models can not take DICOM format images as input, so we converted DICOM format to JPEG format. The image resolution for each MRI image is  $256 * 256$ , see Figure 3.1.



*Figure 3.1.* On the top row, images show the 3rd, 32th and 50th slice of a ASD patient brain MRI. On the bottom row, images show the 80th, 110th and 156th slice of the same ASD patient brain MRI.

The original MRI images have some noises (black areas) around the brain tissues, which can reduce the prediction accuracy for ML models. Thus, we cropped a  $128 * 128$  area from the original image. The reason we used  $128 * 128$  resolution is because our models can get a  $7 * 7$  feature map after subsampling without losing any information. Although there are lots of ways to select a  $128 * 128$  area from a  $256 * 256$  image, we used the central part of the original image. See Figure 3.2.

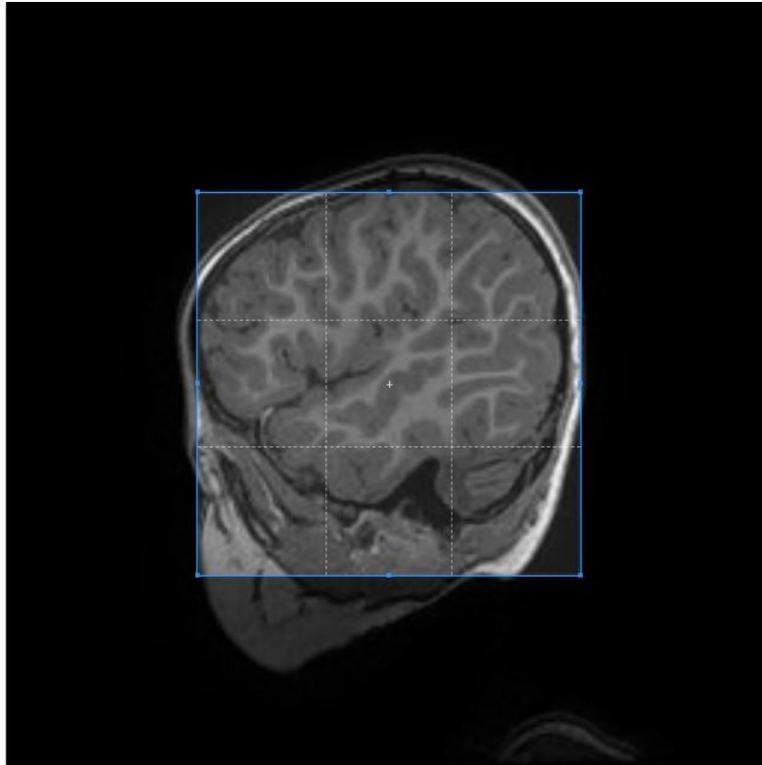


Figure 3.2. The grey box shows the cropping boundary. The coordinator for the top left vertex is (64, 64).

In this study, we can describe the data by using formula:

$$data = w \times h \times d \times c \quad (3.1)$$

In the formula, w is "width", h is "height", d is "depth" and c is "channel". The "width" and "height" is the resolution for the image, which is 128 x 128. The "depth" for this study is set to 156, because each object has 156 slices of MRI images. The "channel" for this study is set to 3, because each MRI image has a red channel, green channel and blue channel.

### 3.4 ML Model and Algorithms

We constructed three Machine Learning models in this study, which are Conv3D, VGG-16 and 3D ResNet.

### 3.4.1 Conv3D

The first model we tried is Conv3D. The model we constructed has 8 convolution layers, 5 max-pooling layers and three fully connected layers. The input for the model is a tensor with the size of  $(128 \times 128 \times 156 \times 3)$ . For each convolution layer, we used four parameters, which are the number of input and output features, kernel size and padding. The convolution function can be written in python as:

```
1 import torch.nn as nn
2 self.conv1 = nn.Conv3d(in_features, out_features, kernel_size, padding)
```

Listing 3.1: Convolution function.

For each max-pooling layer, we used two parameters, which are the kernel size and stride. The max-pooling function can be written in python as:

```
1 import torch.nn as nn
2 self.pool1 = nn.MaxPool3d(kernel_size, stride)
```

Listing 3.2: Max-pooling function.

For each fully-connected layer, we used two parameters, which are amount of input features and amount of output features. The fully connected function can be written in python as:

```
1 import torch.nn as nn
2 self.fc1 = nn.Linear(in_features, out_features)
```

Listing 3.3: Fully connected function.

In this study, we only have around 80 training samples, but fully connected layer has 4096 features. In order to avoid overfitting problem, we added a "dropout" function after fully connected layers and set  $p = 0.5$ , which means half of hidden neurons was deleted temporally for each epoch. The dropout function can be written in python as:

```
1 import torch.nn as nn
2 self.dropout = nn.Dropout(p=0.5)
```

Listing 3.4: Dropout function.

The activation function we used for this model is ReLu, which set all negative numbers to 0 and keep positive numbers. The formula of ReLu function is:

$$ReLU = \begin{cases} x & , if x > 0 \\ 0 & , if x < 0 \end{cases} \quad (3.2)$$

We set the “inplace” value to ”True” for the ReLU function, which could save memories.

The ReLU function can be written in python as:

```
1 import torch.nn as nn
2 self.relu = nn.ReLU(inplace=True)
```

Listing 3.5: ReLU function.

Although this study has two output classes, which are ASD and TD, we want the model to be capable of classifying more than two classes in the future , such as HFA class and LFA class.

Thus, we used softmax function as the classifier. The score function of softmax can be written as:

$$f(x_i; W) = W_{x_i} \quad (3.3)$$

The loss function of softmax is CrossEntropy, which can be written as:

$$L_i = -\log\left(\frac{e^{f_{y_i}}}{\sum_j e^{f_j}}\right) \quad (3.4)$$

This study wants to solve a binary classification problem, so we can simplify the CrossEntropy formula as:

$$-(y * \log(p) + (1 - y) * \log(1 - p)) \quad (3.5)$$

The softmax function formula is:

$$\sigma(z_j) = \frac{e^{z_j}}{\sum_k e^{z_k}} \quad (3.6)$$

The softmax function can be written in python as:

```
1 import torch.nn as nn
2 self.softmax = nn.softmax()
```

Listing 3.6: Softmax function.

The Conv3D model structure is shown in Table 3.6.

Table 3.6. *The Conv3D model.*

input(128 x 128 x 156 RGB volume)
conv-1: $3 \times 3 \times 3 - 64$
maxpool3D-1: kernel(1,2,2)
conv-2: $3 \times 3 \times 3 - 64$
maxpool3D-2, kernel(2,2,2)
conv-3a: $3 \times 3 \times 3 - 128$
conv-3b: $3 \times 3 \times 3 - 128$
maxpool3D-3: kernel(2,2,2)
conv-4a: $3 \times 3 \times 3 - 256$
conv-4b: $3 \times 3 \times 3 - 256$
maxpool3D-4: kernel(2,2,2)
conv-5a: $3 \times 3 \times 3 - 256$
conv-5b: $3 \times 3 \times 3 - 256$
maxpool3D-5: kernel(2,2,2)
FC-6: 4096
FC-7: 4096
FC-8: 4096
dropout: p=0.5
relu
softmax

### 3.4.2 VGG-16

The second model we tried is VGG-16, but we removed some layers in order to decrease parameter sizes. The model we constructed has 10 convolution layers, 5 max-pooling layers and two fully connected layers. The main advantage of VGG is replacing large kernels with several smaller kernels, which can decrease parameter sizes significantly. Thus, we grouped 2 convolution layers as one layer, and followed by a max-pooling layer. Each convolution layer have a  $(3 \times 3 \times 3)$  kernel. Although the accuracy of this model is not good, we showed this model runs faster than Conv3D model, which was discussed in Chapter 4. The input for the model is a tensor with the size of  $(128 \times 128 \times 156 \times 3)$ . All parameters used for convolution layer, max-pooling layer, fully connected layer, dropout function and ReLu function are the same Conv3D model. However, the VGG model has different input features and output features for each layer. The model structure is shown in Table3.7.

Table 3.7. *The 3D-VGG model.*

input(128 x 128 x 156 RGB volume)
conv-1a: $3 \times 3 \times 3 - 16$
conv-1b: $3 \times 3 \times 3 - 16$
maxpool3D-1: kernel(1,2,2)
conv-2a: $3 \times 3 \times 3 - 32$
conv-2b: $3 \times 3 \times 3 - 32$
maxpool3D-2: kernel(2,2,2)
conv-3a: $3 \times 3 \times 3 - 32$
conv-3b: $3 \times 3 \times 3 - 32$
maxpool3D-3: kernel(2,2,2)
conv-4a: $3 \times 3 \times 3 - 64$
conv-4b: $3 \times 3 \times 3 - 64$
maxpool3D-4: kernel(2,2,2)
conv-5a: $3 \times 3 \times 3 - 128$
conv-5b: $3 \times 3 \times 3 - 128$
maxpool3D-5: kernel(2,2,2)
FC-6: 256
FC-7: 256
dropout: p=0.5
relu

### 3.4.3 3D ResNet

In the 3D ResNet model, we did some modifications for building blocks (He et al., 2016). Instead of using two  $3 \times 3$  kernels for building blocks, we used two  $3 \times 3 \times 3$  volume kernels to learn spatiotemporal features from different MRI layers. We still keep the shortcut feature to avoid vanishing gradients. Our building block structure is shown as Figure 3.3.

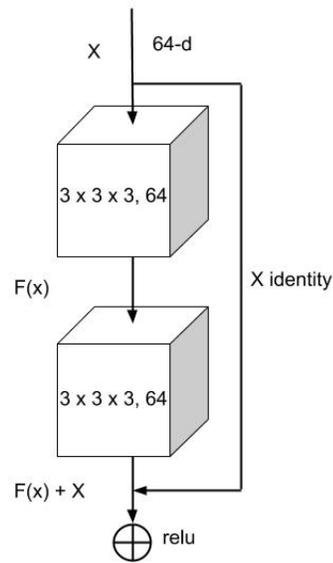


Figure 3.3. There are two  $(3 \times 3 \times 3, 64)$  kernels in each block.

According to our building block,  $x$  represents the input,  $F(x)$  represents the residue function,  $(F(x) + x)$  represents the output for the block.

The 3D ResNet model has four building blocks. We first used one convolution layer for subsampling. Then, we added one max-pooling layer and four building blocks to extract features. Next, we used one average-pooling function, one fully connected layer and a softmax function to classify ASD and TD group. The average-pooling function can be written in python as:

```
1 import torch.nn as nn
2 self.avgpool = nn.AdaptiveAvgPool3d((1,1,1))
```

Listing 3.7: Average-pooling function.

Instead of implementing the convolution layer itself, we implemented convolution blocks for the 3D ResNet model. Four parameters are used when we build the model, which are block structure, layer index, the number of filters and stride. Therefore, the model has four convolution block layers and each layer has one block. These four convolution block layers can be written in python as:

```

1 block_layer_1=self.make_layer(block,layer[0],64, stride=1)
2 block_layer_2=self.make_layer(block,layer[1],128, stride=2)
3 block_layer_3=self.make_layer(block,layer[2],256, stride=2)
4 block_layer_4=self.make_layer(block,layer[3],512, stride=2)

```

Listing 3.8: Four convolution block layers.

The 3D ResNet model structure is shown in Table 3.8.

Table 3.8. *The 3D ResNet model.*

input(128 x 128 x 156 RGB volume)		
layer name	output size	10-layer
conv-1	64	7 x 7 x 7, 64, stride 2, padding 3
conv2_x	64	3 x 3 x 3 maxpool, stride 2
		3 x 3 x 3, 64 3 x 3 x 3, 64
conv3_x	128	3 x 3 x 3, 64 3 x 3 x 3, 64
		3 x 3 x 3, 64 3 x 3 x 3, 64
conv4_x	256	3 x 3 x 3, 64 3 x 3 x 3, 64
		3 x 3 x 3, 64 3 x 3 x 3, 64
conv5_x	512	3 x 3 x 3, 64 3 x 3 x 3, 64
		3 x 3 x 3, 64 3 x 3 x 3, 64
	7 x 7	average pool(1,1,1), FC: 512, softmax

### 3.4.4 Parameters Configuration

For each model, we used Xavier initialization to initialize weights for the network. The Xavier initialization can be represented by formula:

$$std = gain \times \sqrt{\frac{2}{fan\_in + fan\_out}} \quad (3.7)$$

In this equation, "fan\_in" is incoming network connections, and "fan\_out" is outgoing network connections from that layer. The Xavier initialization can be written in python as:

```
1 import torch.nn as nn
2 torch.nn.init.xavier_normal_(tensor, gain=1)
```

Listing 3.9: Xavier initialization.

In addition, we used batch size equal to 8, because we want to fully use the GPU's capability. We also tried to set batch size to 16, but it caused memory overflow.

Learning rate is a crucial hyperparameter of ML models. Learning rate represents how fast the ML model has adapted the problem and it should be set properly. The model can converge too fast with a large learning rate, leading to a suboptimal solution rather than the most optimal solution. likewise, the model can get stuck in the optimization process with very small learning rate, because the update step is too small to get the most optimal solution. This study utilized adam optimizer to update learning rate.

We do not have a lot of data for the study. Therefore, we used 20 epochs in order to avoid overfitting problem. We neither want model converge too fast nor too slow, so learning rate is set to  $10^{-3}$  at the initialization stage. It keeps the same value for the first and the second epochs. Then, we set the learning rate to  $10^{-4}$  from the 3<sup>rd</sup> epoch to the 12<sup>th</sup> epoch. Finally, we updated the learning rate to  $10^{-5}$  from 13<sup>th</sup> to 20<sup>th</sup> epoch. There were 20 epochs in total.

### 3.5 Variables

The independent variable for this study is different ML models, including Conv3D, VGG and 3D ResNet. The dependent variables for this research is the accuracy of ASD prediction and the speed of diagnosing processes.

### 3.6 Testing Procedures

The testing procedures can be divided into three parts. The first part is calculating and comparing the performance among three ML models, in terms of prediction accuracy and speed. The second part is comparing the performance between ML models and traditional behavioral tests.

For the first part of testing, we used accuracy, precision, recall and F1 score to evaluate the performance of models. And we used time cost to evaluate the speed. Accuracy is the correct prediction ratio of testing data. The formula to calculate accuracy is:

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative} \quad (3.8)$$

Recall can be interpreted as how many relevant objects are selected. In simple words, recall represents whether the model finds all true ASD patients in the testing dataset. The range of recall is  $[0, 1]$ , the higher value of recall means the better performance to find true ASD patients. The recall is calculated by the following formula:

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (3.9)$$

Precision can be interpreted as how many selected objects are relevant. In simple words, precision represents whether the predicted ASD objects are truly ASD patients. The precision is calculated by the following formula:

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (3.10)$$

F1 is used to balance precision and recall. In this study, we want all true ASD objects to be found, even though some TD objects may be misclassified. In simple words, we do not want to miss any true ASD patients. Thus, getting a relatively high recall value is our priority. The  $F_1$  score can be calculated as:

$$F_1 = 2 \times \frac{precision \times recall}{precision + recall} \quad (3.11)$$

In order to count the speed for each model, we added a timer to record the time consumption for each object. Then we used average time consumption to represent the speed of models. The average time is calculated by the following formula:

$$avg\_time = \frac{\sum_{i=1}^n t_i}{n} \quad (3.12)$$

In the fomrmula above, "t" represents the time used to make predictions and "n" represents the number of objects.

The second step of the testing procedure is comparing the performance between ML models and traditional ASD behavioral tests. We used the best diagnosis accuracy and the shortest time consumption for behavioral tests as the baseline.

### 3.7 Analysis

The final result was measured by five parameters, which are accuracy, precision, recall,  $F_1$  and average time. All parameters were recorded and printed on the terminal for each model. We compared the ML method performance with the best ASD behavioral test performance to confirm or reject the null hypothesis.

### 3.8 Summary

In this chapter, we first proposed the hypothesis of this study. Then, both hardware and software configuration were indicated. Next, we introduced our dataset and three Machine Learning models. Finally, we listed variables and talked about the testing procedure and analysis. In the next chapter, we discussed the experiment results.

## CHAPTER 4. RESULTS

This chapter includes the experiment result for three machine learning models and the performance comparison. Also, a performance comparison between ASD behavioral tests and ML models was provided.

In this study, the "true positive (TP)", "false positive (FP)", "false negative (FN)" and "true negative (TN)" are defined in Table 4.1.

Table 4.1. *The confusion matrix.*

	Condition positive	Condition negative
Predicted condition positive	TP: ASD object, prediction is ASD	FP: TD object, prediction is ASD
Predicted condition negative	FN: ASD object, prediction is TD	TN: TD object, prediction is TD

### 4.1 Conv3D Results

The result for Conv3D model is shown in Table 4.2

Table 4.2. *The best result for Conv3D.*

Model: Conv3D	
Accuracy	56.5217
Precision	0.5652
Recall	1.0000
F.1 Score	0.7222
Average Time	0.4608

## 4.2 VGG Results

The result for VGG model is shown in Table 4.3.

Table 4.3. *The best result for VGG.*

Model: VGG	
Accuracy	65.2174
Precision	0.9375
Recall	0.6818
F_1 Score	0.7895
Average Time	0.4282

## 4.3 3D ResNet Results

The result for the 3D ResNet model is shown in Table 4.4.

Table 4.4. *The best result for 3D ResNet.*

Model: 3D ResNet	
Accuracy	86.9565
Precision	0.9091
Recall	0.9524
F_1 Score	0.9302
Average Time	0.3893

## 4.4 Results Comparison Among Different Models

The performance comparison among three ML models is shown in Figure 4.1.

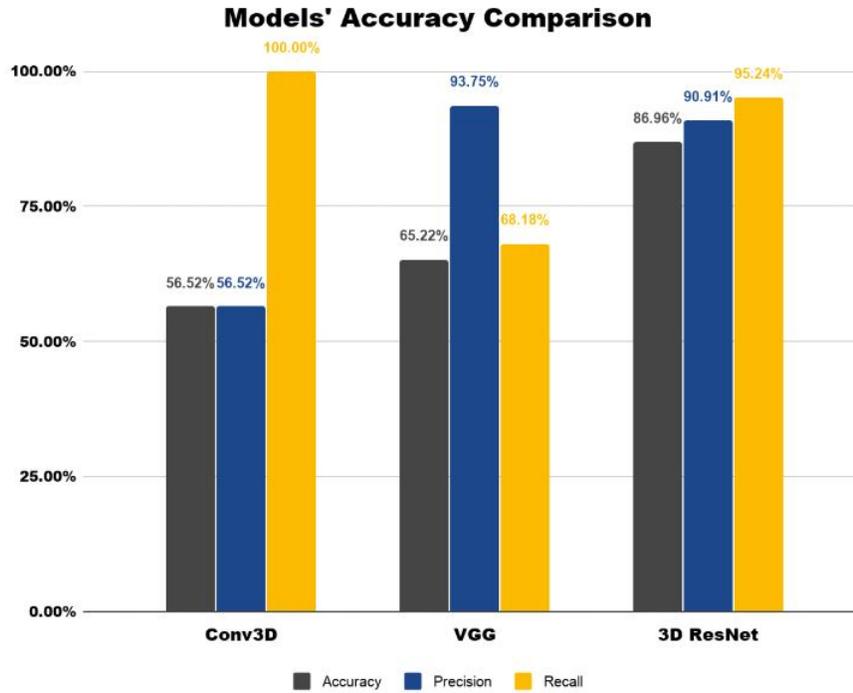


Figure 4.1. Models' Performance Comparison.

The figure above showed that Conv3D has the highest recall score of 100%, which means it did not miss any true ASD object. However, the accuracy and precision of Conv3D is the lowest among three models, which means it misclassified lots of TD objects to ASD class. The VGG model has the highest precision score of 93.75%, which means most of objects in ASD class is valid. But VGG misses some true ASD objects and the accuracy is also relatively low. 3D ResNet has the best performance overall. It has the highest accuracy of 86.96%. The recall score of 3D ResNet is 95.24%, which covered most of true ASD objects. The precision score of 3D ResNet is 90.91%, which means most of objects in ASD class is valid.

The speed comparison among three ML models is shown in Figure 4.2.

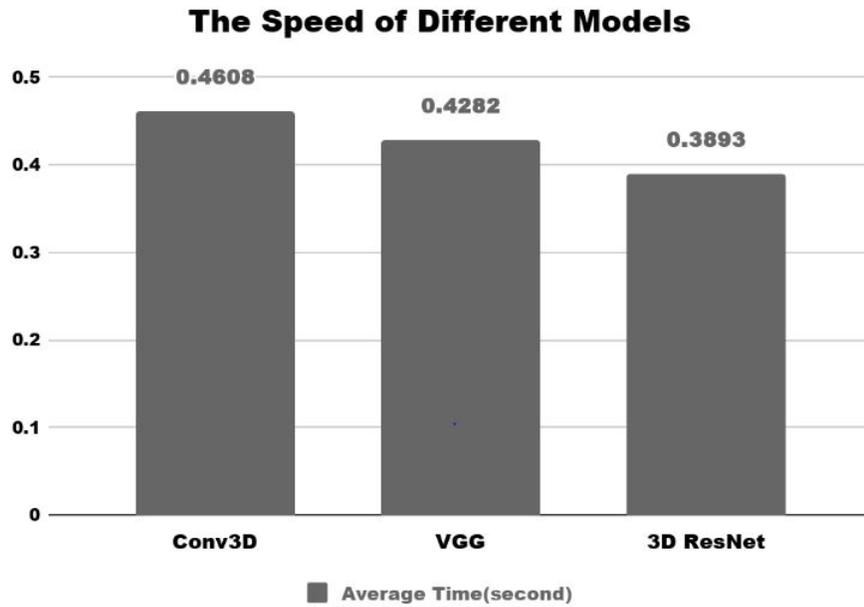


Figure 4.2. Models' Speed Comparison.

The figure above showed that 3D ResNet has the fastest speed among three models, which can make predictions in 0.39 second for each object. The speed differences among three models is not huge, and all of them are much faster than ASD behavioral tests.

#### 4.5 Results Compared with Behavioral Screening Tools

The best performed ML model for this study is 3D ResNet, so we compared traditional behavioral tests with 3D ResNet. See Figure 4.3 for performance comparison.

### Accuracy Comparison of 3D ResNet and Behavioral Test

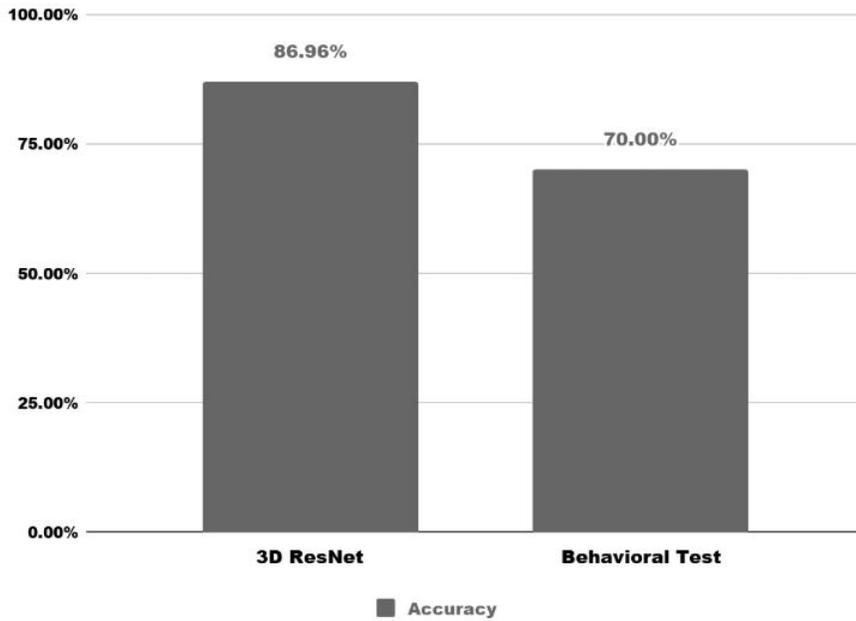


Figure 4.3. Accuracy Comparison: ML Method vs. Behavioral Test

The figure above shows that the ASD prediction accuracy of 3D ResNet is 86.96%, which is better than the highest behavioral test accuracy with 70%.

The speed comparison between 3D ResNet and behavioral tests is shown in Figure 4.4.

### Speed Comparison of 3D ResNet and Behavioral Test

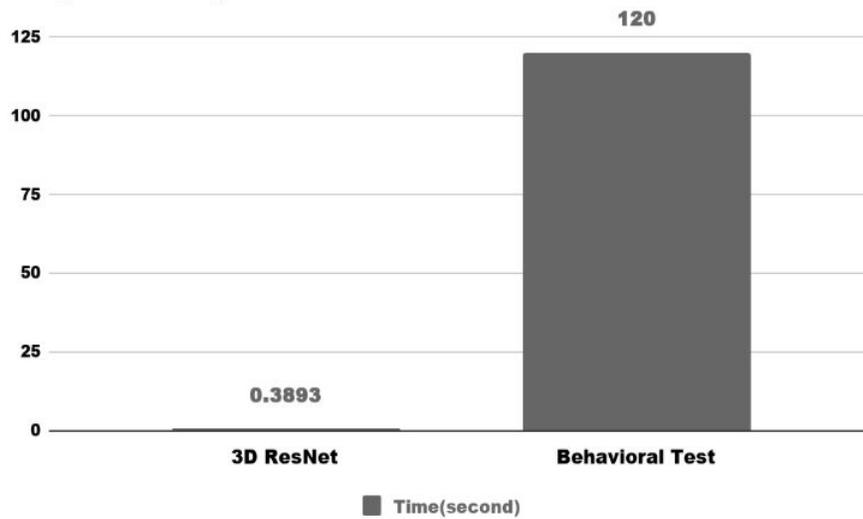


Figure 4.4. Speed Comparison: ML Method vs. Behavioral Test Speed

The figure above shows that the fastest ASD behavioral test needs 120 *second* for each prediction, which is 308 times longer than 3D ResNet.

#### 4.6 Discussion

In this study, we can conclude that not every ML model has good performance when dealing with structural MRI data. 3D ResNet is not only the best model among three models, but also better than behavioral tests in terms of accuracy and speed. 3D ResNet has 17% higher ASD prediction accuracy and 308*times* faster speed than traditional ASD behavioral tests. In addition, we found that the Conv3D model predicted all objects in the testing dataset are ASD, which means it did not learn features correctly or it was incapable to deal with structural MRI. We tried to modify the structure and parameters for Conv3D as the future work.

#### 4.7 Summary

In this chapter, we showed the experiment result for three different ML models and compared their performance and speed. We also compared ML models' performance with traditional behavioral tests. We concluded this study in the next chapter.

## CHAPTER 5. CONCLUSION

This thesis investigated three Machine Learning models for ASD detection and compared ML models with traditional ASD behavioral tests. According to the result from chapter 4, two null hypotheses are rejected. At the same time, two alternative hypotheses are accepted. which are:

- The use of ML method does improve the accuracy of ASD diagnosis by using MRI images.
- The use of ML method does increase the speed of ASD diagnosis by using MRI images.

We can conclude experiment results of the study as follows:

- ML models are able to detect ASD by using structural MRI with faster speed and higher accuracy than behavioral tests.
- ML models' result is unexplainable in both pathology and behavioral observations. ML models served as a "black box", which is not satisfied by most medical professionals. In clinical practices, neither patients nor professionals rely on predictions that are made by computer. Therefore, it is crucial to find biomarkers and make the model explainable.
- Structural MRI has too many noises for ML models, so it is very important to design a data preparation process.
- ML models can only assist medical professionals to make diagnosis at the current stage, rather than make diagnosis by themselves.
- A much bigger dataset is necessary to prove the reliability and validity of our models.

## 5.1 Future Works

The future work of this study will mainly focus on finding biomarkers from the ML models. In this study, our ML models served as a blackbox, which means they can only output unexplainable predictions. This could be tolerated in some field, but not in clinical fields. Every medical professional has to explain the pathology and causes for their patients, rather than providing a simple prediction score. So, it is very important to capacitate our ML models with explainable results.

Another future work is to test our ML models with bigger dataset, such as BRIDE. Even though our models showed good prediction accuracy with our private dataset, it does not mean that our models can make accurate predictions on other dataset. The first reason is that our dataset is relatively small compared with other public datasets. Even though we used cross-validation to train our models as much as we can, we used only 22 objects as our testing dataset. Thus, each object in the testing dataset can affect the final prediction a lot. The second reason is that our private dataset is collected by a group of professional radiologists in the same experiment environment, so the quality of our dataset is higher than public dataset. On the contrary, public dataset is collected by different research groups and hospitals by different MRI machines. In addition, the age range could be much larger than our private dataset. In this situation, our models may not be able to achieve the same high prediction accuracy as we did. The biggest advantage of a public dataset is size, which is much larger than our dataset. Eventually, we want to improve our models and fine-tune all parameters to get a higher ASD diagnosis accuracy for the most of brain MRI dataset.

## REFERENCES

- Albelwi, S., & Mahmood, A. (2017). A framework for designing the architectures of deep convolutional neural networks. *Entropy*, *19*(6), 242.
- Arnold, L. E., Aman, M. G., Li, X., Butter, E., Humphries, K., Scahill, L., . . . Stigler, K. A. (2012). Research units of pediatric psychopharmacology (rupp) autism network randomized clinical trial of parent training and medication: One-year follow-up. *Journal of the American Academy of Child Adolescent Psychiatry*, *51*(11), 1173–1184.
- Bandim, J. M., Ventura, L. O., Miller, M. T., Almeida, H. C., & Costa, A. E. S. (2003). Autism and möbius sequence: an exploratory study of children in northeastern brazil. *Arquivos de Neuro-psiquiatria*, *61*(2A), 181–185.
- Bernal, J., Kushibar, K., Asfaw, D. S., Valverde, S., Oliver, A., Martí, R., & Lladó, X. (2019). Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: a review. *Artificial intelligence in medicine*, *95*, 64–81.
- Boiman, O., & Irani, M. (2007). Detecting irregularities in images and in video. *International journal of computer vision*, *74*(1), 17–31.
- Boissoneault, J., Sevel, L., Letzen, J., Robinson, M., & Staud, R. (2017). Biomarkers for musculoskeletal pain conditions: Use of brain imaging and machine learning. *Current Rheumatology Reports*, *19*(1), 1–9.
- Buescher, A. V. S., Cidav, Z., Knapp, M., & Mandell, D. S. (2014, 08). Costs of Autism Spectrum Disorders in the United Kingdom and the United States . *JAMA Pediatrics*, *168*(8), 721-728. Retrieved from <https://doi.org/10.1001/jamapediatrics.2014.210>  
doi: 10.1001/jamapediatrics.2014.210
- Centers for Disease Control and Prevention. (2019). *Screening and diagnosis of autism spectrum disorder*. Retrieved from <https://www.cdc.gov/ncbddd/autism/signs.html>
- Centers for Disease Control and Prevention. (2020a). *Data statistics on autism spectrum disorder*. Retrieved from <https://www.cdc.gov/ncbddd/autism/data.html>
- Centers for Disease Control and Prevention. (2020b). *Screening and diagnosis of autism spectrum disorder*. Retrieved from <https://www.cdc.gov/ncbddd/autism/screening.html>
- Centers for Disease Control and Prevention. (2020c). *What is autism spectrum disorder?* Retrieved from <https://www.cdc.gov/ncbddd/autism/facts.html>

- Chen, R., Jiao, Y., & Herskovits, E. H. (2011). Structural mri in autism spectrum disorder. *Pediatric research*, 69(8), 63–68.
- Ciregan, D., Meier, U., & Schmidhuber, J. (2012). Multi-column deep neural networks for image classification. In *2012 ieee conference on computer vision and pattern recognition* (pp. 3642–3649).
- Committee on Children with Disabilities and others. (2001). Developmental surveillance and screening of infants and young children. *Pediatrics*, 108(1), 192–195.
- Di Martino, A., Yan, C.-G., Li, Q., Denio, E., Castellanos, F. X., Alaerts, K., . . . others (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.
- Dobrez, D., Sasso, A. L., Holl, J., Shalowitz, M., Leon, S., & Budetti, P. (2001). Estimating the cost of developmental and behavioral screening of preschool children in general pediatric practice. *Pediatrics*, 108(4), 913–922.
- DSM-IV. (1994). Task force on dsm-iv.(1994). *Diagnostic and statistical manual of mental disorders: DSM-IV*.
- Fischl, B. (2012). Freesurfer. *Neuroimage*, 62(2), 774–781.
- Garstang, J., & Wallis, M. (2006). Randomized controlled trial of melatonin for children with autistic spectrum disorders and sleep problems. *Child: care, health and development*, 32(5), 585–589.
- Giedd, J. N. (2004). Structural magnetic resonance imaging of the adolescent brain. *Annals of the New York Academy of Sciences*, 1021, 77.
- Glascoc, F. P. (1998). *Collaborating with parents: Using parents' evaluation of developmental status to detect and address developmental and behavioral problems*. Ellsworth & Vandermeer Press.
- Hallmayer, J., Cleveland, S., Torres, A., Phillips, J., Cohen, B., Torigoe, T., . . . Risch, N. (2011, 11). Genetic Heritability and Shared Environmental Factors Among Twin Pairs With Autism. *Archives of General Psychiatry*, 68(11), 1095-1102. Retrieved from <https://doi.org/10.1001/archgenpsychiatry.2011.76> doi: 10.1001/archgenpsychiatry.2011.76
- Hazlett, H. C., Gu, H., Munsell, B. C., Kim, S. H., Styner, M., Wolff, J. J., . . . others (2017). Early brain development in infants at high risk for autism spectrum disorder. *Nature*, 542(7641), 348–351.

- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).
- 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*, 16–23.
- Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700–4708).
- Hutsler, J. J., Love, T., & Zhang, H. (2007). Histological and magnetic resonance imaging assessment of cortical layering and thickness in autism spectrum disorders. *Biological Psychiatry, 61*(4), 449 - 457. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0006322306001478> (Advances in Understanding and Treating Autism Spectrum Disorders) doi: <https://doi.org/10.1016/j.biopsych.2006.01.015>
- Itzhak, E. B., Lahat, E., & Zachor, D. A. (2011). Advanced parental ages and low birth weight in autism spectrum disorders—rates and effect on functioning. *Research in developmental disabilities, 32*(5), 1776–1781.
- Ivakhnenko, A. G., & Lapa, V. G. (1966). *Cybernetic predicting devices* (Tech. Rep.). PURDUE UNIV LAFAYETTE IND SCHOOL OF ELECTRICAL ENGINEERING.
- Ivakhnenko, A. G., & Lapa, V. G. (1967). *Cybernetics and forecasting techniques*.
- Jain, V., Bollmann, B., Richardson, M., Berger, D. R., Helmstaedter, M. N., Briggman, K. L., . . . others (2010). Boundary learning by optimization with topological constraints. In *2010 IEEE computer society conference on computer vision and pattern recognition* (pp. 2488–2495).
- Ji, S., Xu, W., Yang, M., & Yu, K. (2012). 3d convolutional neural networks for human action recognition. *IEEE transactions on pattern analysis and machine intelligence, 35*(1), 221–231.
- Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. *Computational and structural biotechnology journal, 13*, 8–17.

- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097–1105).
- Laptev, I. (2005). On space-time interest points. *International journal of computer vision*, *64*(2-3), 107–123.
- LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural computation*, *1*(4), 541–551.
- Li, X., Dvornek, N. C., Zhuang, J., Ventola, P., & Duncan, J. S. (2018). Brain biomarker interpretation in asd using deep learning and fmri. In *International conference on medical image computing and computer-assisted intervention* (pp. 206–214).
- Lin, W., Tong, T., Gao, Q., Guo, D., Du, X., Yang, Y., ... others (2018). Convolutional neural networks-based mri image analysis for the alzheimer's disease prediction from mild cognitive impairment. *Frontiers in neuroscience*, *12*, 777.
- Liu, W., Yu, X., Raj, B., Yi, L., Zou, X., & Li, M. (2015). Efficient autism spectrum disorder prediction with eye movement: A machine learning framework. In *2015 international conference on affective computing and intelligent interaction (acii)* (pp. 649–655).
- Loomes, R., Hull, L., & Mandy, W. P. L. (2017). What is the male-to-female ratio in autism spectrum disorder? a systematic review and meta-analysis. *Journal of the American Academy of Child & Adolescent Psychiatry*, *56*(6), 466–474.
- Lord, C., Risi, S., DiLavore, P. S., Shulman, C., Thurm, A., & Pickles, A. (2006). Autism from 2 to 9 years of age. *Archives of general psychiatry*, *63*(6), 694–701.
- Lundervold, A. S., & Lundervold, A. (2019). An overview of deep learning in medical imaging focusing on mri. *Zeitschrift für Medizinische Physik*, *29*(2), 102–127.
- Maenner MJ, B. J. e. a., Shaw KA. (2020). Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2016. *MMWR Surveill Summ* 2020. doi: 10.15585/mmwr.ss6904a1
- Mann, J. R., McDermott, S., Bao, H., Hardin, J., & Gregg, A. (2010). Pre-eclampsia, birth weight, and autism spectrum disorders. *Journal of autism and developmental disorders*, *40*(5), 548–554.

- Markram, H., Rinaldi, T., & Markram, K. (2007). The intense world syndrome-an alternative hypothesis for autism. *Frontiers in neuroscience*, *1*, 6.
- Martin-Isla, C., Campello, V., Izquierdo, C., Raisi-Estabragh, Z., Baeßler, B., Petersen, S., & Lekadir, K. (2020). Image-based cardiac diagnosis with machine learning: A review. *Frontiers in Cardiovascular Medicine*, *7*.
- Miller, A., Alston, R., & Corsellis, J. (1980). Variation with age in the volumes of grey and white matter in the cerebral hemispheres of man: measurements with an image analyser. *Neuropathology and applied neurobiology*, *6*(2), 119–132.
- Mittal, S. (2018). A survey of fpga-based accelerators for convolutional neural networks. *Neural computing and applications*, 1–31.
- Moore, S., Turnpenny, P., Quinn, A., Glover, S., Lloyd, D., Montgomery, T., & Dean, J. (2000). A clinical study of 57 children with fetal anticonvulsant syndromes. *Journal of medical genetics*, *37*(7), 489–497.
- Nanson, J. (1992). Autism in fetal alcohol syndrome: a report of six cases. *Alcoholism: Clinical and Experimental Research*, *16*(3), 558–565.
- Orrù, G., Pettersson-Yeo, W., Marquand, A. F., Sartori, G., & Mechelli, A. (2012). Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review. *Neuroscience Biobehavioral Reviews*, *36*(4), 1140 - 1152. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0149763412000139> doi: <https://doi.org/10.1016/j.neubiorev.2012.01.004>
- Over, P., Awad, G., Michel, M., Fiscus, J., Sanders, G., Shaw, B., ... Quéot, G. (2013). Trecvid 2012-an overview of the goals, tasks, data, evaluation mechanisms and metrics.
- Rasalam, A., Hailey, H., Williams, J. H. G., Moore, S., Turnpenny, P., Lloyd, D. J., & Dean, J. C. (2005). Characteristics of fetal anticonvulsant syndrome associated autistic disorder. *Developmental medicine and child neurology*, *47*(8), 551–555.
- Retson, T., A., Besser, A., H., Sall, S., H., Golden, D., H., & Hsiao, A., H. (2019). Machine learning and deep neural networks in thoracic and cardiovascular imaging. *Journal of Thoracic Imaging*, *34*(3), 192–201.
- Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, *3*(3), 210–229.
- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, *61*, 85–117.

- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Singh, K., Connors, S. L., Macklin, E. A., Smith, K. D., Fahey, J. W., Talalay, P., & Zimmerman, A. W. (2014). Sulforaphane treatment of autism spectrum disorder (asd). *Proceedings of the National Academy of Sciences*, *111*(43), 15550–15555. Retrieved from <https://www.pnas.org/content/111/43/15550> doi: 10.1073/pnas.1416940111
- Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., ... others (2004). Advances in functional and structural mr image analysis and implementation as fsl. *Neuroimage*, *23*, S208–S219.
- Strömmland, K., Nordin, V., Miller, M., Akerström, B., & Gillberg, C. (1994). Autism in thalidomide embryopathy: a population study. *Developmental Medicine & Child Neurology*, *36*(4), 351–356.
- Tran, D., Bourdev, L., Fergus, R., Torresani, L., & Paluri, M. (2015). Learning spatiotemporal features with 3d convolutional networks. In *Proceedings of the ieee international conference on computer vision* (pp. 4489–4497).
- Turaga, S. C., Murray, J. F., Jain, V., Roth, F., Helmstaedter, M., Briggman, K., ... Seung, H. S. (2010). Convolutional networks can learn to generate affinity graphs for image segmentation. *Neural computation*, *22*(2), 511–538.
- Volkmar, F. R., Lord, C., Bailey, A., Schultz, R. T., & Klin, A. (2004). Autism and pervasive developmental disorders. *Journal of child psychology and psychiatry*, *45*(1), 135–170.
- Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). Aggregated residual transformations for deep neural networks. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 1492–1500).
- Zhang, W., et al. (1988). Shift-invariant pattern recognition neural network and its optical architecture. In *Proceedings of annual conference of the japan society of applied physics*.