Improving Autism Diagnosis in India

Applying Machine Learning and Neural Networks for effective prediction

Project Author

Samir Agarwala
Student, The Cathedral and John Connon School, Mumbai

Abstract

Advances in medicine have helped improve effective diagnosis of autism spectrum disorder (ASD) over the years, but still in majority of the cases the diagnosis cannot be made early enough to drastically improve the prognosis of the disorder. Current diagnosis methods are based on a set of behavioral criterion and not on any biological indicator and thus are subjective, error prone and time consuming. Machine Learning has the potential to overcome these deficits and predict autism accurately. Accumulation of 3D fMRI brain scans and clinical phenotype data of ASD and non-ASD patients have led to increased possibility of using machine learning for diagnosis purposes. This paper applies and evaluates deep learning models to predict ASD and learn important features from the fMRI scans. The methodology includes a Convolutional AutoEncoder, and a Convolution neural network to learn features and make predictions. Regression analysis is also applied on clinical phenotype data of ASD and non-ASD patients. It is seen that our Deep Models on fMRI data are not predictive of ASD but simple regression analysis classifies correctly with an accuracy of 70% showing that normal features collected in clinics can be used to diagnose autism accurately and quickly.

Project Code

`https://github.com/samiragarwala/autism_prediction`

Overleaf was used to typeset this document. Credit to Autism Brain Imaging Data Exchange (ABIDE) for granting me access to their dataset. R was used to retrieve and preprocess ABIDE datasets. Python was used to implement the neural networks and process data. Keras was used as the neural network package in this project. Credit to Keylekan from Aalborg University for sharing their project template.

Contents

Preface

Autism Spectrum Disorder is a major health disorder affecting millions of people in India. Due to inadequate facilities and lack of trained medical personnel, there is a huge need to work on improving autism diagnosis in India. Apart from this, there is need for a method to diagnose autism early as early intervention can improve the prognosis of the disorder.

Through my personal experiences with autistic children and also my interaction at Ummeed Child Development Center, I gained a clear focus in trying to improve autism diagnosis in India. I was moved by my interaction with the team member and what I observed at the center, and felt that autism diagnosis would be a good place I could apply my machine learning knowledge to try and solve a major disorder affecting millions of people in India and the world.

Samir Agarwala
<samir.aggarwala@gmail.com>

Chapter 1

Introduction

Autism is a highly prevalent neuro-development disorder across the world. In developing countries such as India, proper large-scale diagnostic and treatment facilities for such disorder are not present. This makes the accurate diagnosis of autism an important problem to address since currently diagnosis of autism is based on a behavioral criteria and not on any bio-markers. A medical professional assesses a patient for the disorder using the prescribed norms, but in the process a substantial amount of misdiagnoses and late diagnoses occur. This leads to most autism cases being diagnosed after the age of 5, whereas early intervention could significantly improve the autistic person's life. Due to this, it is important to find a way to accurately detect autism using some bio-markers.

Considering the difficulty of autism diagnosis, it is important to find a way to solve it. A step in the right direction would be to apply machine learning techniques such as deep learning to detect autism using fMRIs and other biological data.

Deep Learning is a branch of machine learning which is used to model high level abstractions in data. These models are some of the most accurate machine learning algorithms in modern times. These models are applied everywhere- from search engines to recommender systems to cancer detection. Due to this, they have huge potential in accurately diagnosing autism. Hence, experimentation must be done with these models to find a way to detect autism accurately.

In this report, we examine the use of deep learning models to detect autism using fMRI scans, while also using regression analysis on phenotype data of autistic patients for the same purpose. We believe that by addressing such an important challenge through a machine learning approach, we will be able to contribute constructively to research on accurate autism diagnosis.

Chapter 2

Background: Autism

2.1 What is Autism?

Autism Spectrum Disorder (ASD) is a general term used for a group of life-long neuro-development disorders. ASD patients are characterized by lack of social communication and a repetitive pattern of behavior. The following Table 2.1 summarizes the effects of ASD on patients.

Table 2.1: Effects of Autism

Social Communication	**Repetitive Patterns of Behavior**
1. Social and Emotional Reciprocity	1. Repetitive Motor Movements
2. Non-Verbal Communicative Behavior	2. Inflexibility (Adherence to routine)
3. Difficulty in Developing Relations	3. Restricted Fixated Interests
	4. Hypo and Hyper Reactivity to Sensory Input

Let us take a quick look at the Social Communication effects:

- Social and Emotional Reciprocity : ASD patients do not initiate or respond to conversation. They do not share their interests with others and find it tough to continue a conversation.

- Non-Verbal Communicative Behvaviour : ASD patients make abnormal eye contact and body language. They do not gesticulate and have a poorly integrated verbal and non-verbal communication skill set.

- Difficulty in Developing Relations : ASD patients find it very difficult to adjust to the needs of other people and this makes it very hard for them to make friends.

On the basis of the severity of the effects mentioned in Table 2.1 and the amount of support they require , ASD patients are divided into 3 Levels : [16]:

- Level 1 : Require support - Patients are verbal but deficit in social communication. Patients are characterized by inflexibility to changes in their lifestyle and a lack of initiation of social interaction.

- Level 2 : Require substantial support - Patients have a marked deficit in verbal and non-verbal communication. This deficit is apparent even when support is in place for the patient. Their inflexibility interferes in their lifestyle to such an extent that they have difficultly coping with small changes in their life.

- Level 3 : Require very substantial support - Patients are characterized by a very limited initiation of social interaction and severe deficits in verbal and non-verbal communication which considerably hampers their function. They find it extremely difficult coping with change.

2.2 Diagnosis

The current diagnosis for ASD is centered around interactions of patients with medical practitioners , examination of medical and behavioral history, and ruling out of other disorders. Diagnosis is usually done on the basis of behavioral criterion such as the DSM-5. The Diagnostic and Statistical Manual of Mental Disorders (DSM-5) criterion listed by the American Psychological Association is a behavioral criterion summarized in Table 2.1 [12].

2.2.1 Early Diagnosis

It is widely acknowledged that the early diagnosis of ASD leads to early intervention and hence increases the likelihood of an improved prognosis. This is because that the plasticity of the brain is greater at a younger age and hence the autistic children are more likely to respond positively to therapy [21].

ASD affects around 1% [13] of the world's population. The situation of the affected people can be improved a lot by early diagnosis. Currently, more than 50% of children are 5 years or older when they were diagnosed with ASD , while less than 20% of children are diagnosed by age 2 [17]. Diagnosis of autism is possible as early as age 2 when signs of autism start clearly manifesting themselves, but yet diagnosis in more than half the cases does not take place till age 5. Consequently, time which could have been used for effective treatment is lost. This shows that there is a need for an effective diagnosis mechanism which can be used to diagnose autism accurately and early, so as to improve prognosis of affected people.

2.3 Treatment

There is no cure for ASD. However, individual intervention programmes tailored for each patient help improve the condition of the patients. These intervention programmes include both medicinal and behavioral intervention. Medicinal intervention are mainly used to treat problems such as depression, anxiety, hyperactivity, and obsessive-compulsive behaviors which co-exist with ASD in many cases. Behavioral intervention use positive reinforcement and social skills training to improve behavior and communication. Such interventions include: Applied Behavioral Analysis (ABA), Treatment and Education of Autistic and Related Communication Handicapped Children (TEACCH), and sensory integration.[2]

Applied Behavioral Analysis uses a 3 step process [14]:

- <u>Antecedent</u> : Verbal or physics stimulus such as a command or request. This stimulus may be external or internal to the subject.
- <u>Resulting Behavior</u> : The patient's reaction to the antecedent.
- <u>Consequence</u> : It includes the positive reinforcement of the desired behavior for incorrect responses to the antecedent.

Other interventions such as Pivotal Response Treatment, Verbal therapy and the Early Start Denver Model use the three step process described by Applied Behavioral Analysis.[14]

Treatments for ASD are very intensive and involve the patient's family and a team of medical professionals. They aim of the treatment is to improve the patient's quality of life and increase social acceptance. This goal is best achieved in cases of ASD where early intervention is conducted. Statistics show that early diagnosis and intervention reduces the cost of lifelong care of ASD patients by $\frac{2}{3}$ [13]. Additionally, early intervention also leads to a better standard of life among ASD patients. Due to this, more accurate early detection techniques for ASD is one of the most important research topics in the field.

2.4 Current Research

Recent research has been focused on genes and genetic mutations that may contribute to ASD. Studies have uncovered 60 genes that have a more than 90%

chance of contributing to ASD [22]. In another study, researchers used gene expression patterns in blood samples to detect ASD with an accuracy of around 75% [6].

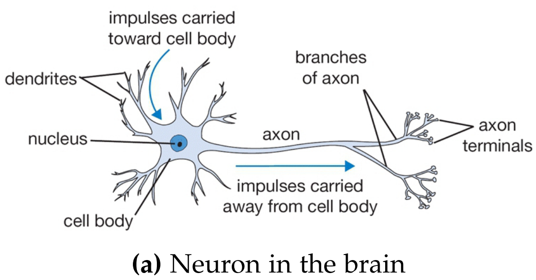
Chapter 3

Background: Neural Networks

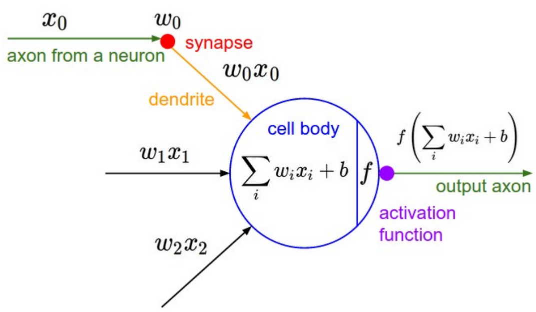
3.1 What are Neural Networks?

Artificial Neural Networks are computing structures modeled to imitate the data processing capability of neural structures in the brain. Traditional machine learning and pattern recognition algorithms are highly linear or mildly non-linear in nature. In an attempt to get high dimensional non-linearities, researchers tried to model the human brain due to its massive computational and comprehension ability. A highly simplified and mathematical model of the brain led to the development of Neural Networks. Each network is built by connecting simple linear computing units called "neurons", in multiple ways, just like inter-neuron connections in the brain.

3.2 Neural Network Architecture



(a) Neuron in the brain



(b) Computational "neuron" model

Figure 3.1: Comparing the biological neuron with the computational neuron, from CS 231N, Stanford University [8]

As explained above, each network, as seen below, consists of computing units called "neurons". Each "hidden unit" takes inputs $x_i \in \mathcal{R}^n$ and then computes the $w^T x$, where $w \in \mathcal{R}^n$ are the weights of the unit. A bias $b \in \mathcal{R}$ is added to get

$$\text{Unit Output } O = w^T x + b$$

3.2.1 Defining the Model

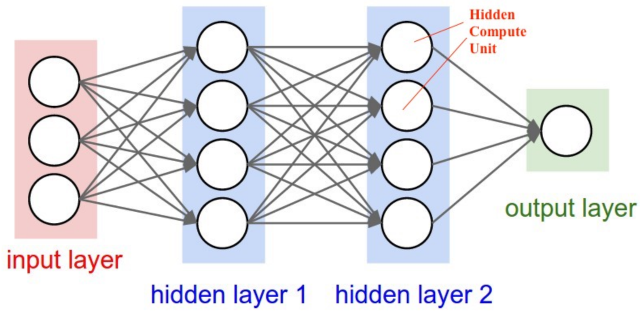


Figure 3.2: Representation of a Neural network, from CS 231N, Stanford University[8]

A neural network is made up of many layers. The first layer is the input layer and the last layer is the output layer. Multiple layers called hidden layers are present between the input and output layer. They are so named, because their values are not observed as outputs. Each hidden unit of a layer is connected to every hidden unit of the next layer, creating a fully connected network.

In a neural network, a non-linear function, called the activation function, is present after each layer. The activation function is usually the sigmoid function or the Rectified Linear Unit (ReLU). If the output of a layer is $o \in \mathcal{R}^n$, then we can represent the output of the activation function, $a \in \mathcal{R}^n$ as

$$a_i = \frac{1}{1 + \exp(-o_i)}, \ \forall \ i$$

for sigmoid and

$$a_i = \max(0, o_i), \ \forall \ i$$

for ReLU.

The output layer is responsible for making the predictions or providing the required "outputs" of the network. The number of units in the output layer is determined by the number of classes being predicted per input. If the network needs to predict a regression value, then we need only 1 output unit. If we need to classify the inputs into Dog, Cat and Cow classes, then we need 3 output units.

Let us look at how a network predicts p inputs $x_i \in \mathcal{R}^n$ using the architecture in Fig 3.2. Our input matrix is now $X \in \mathcal{R}^{p \times n}$. The weights in the input layer need to be $w \in \mathcal{R}^n$ and the bias needs to be $b_r \in \mathcal{R}^n$. Combining all w's and b_r's into matrix form, we get $W_1 \in \mathcal{R}^{n \times 3}$ and $b_1 \in \mathcal{R}^3$. The output of the input layer is $O_1 \in \mathcal{R}^{p \times 3}$. Similarly, the weights for the first hidden layer are $W_2 \in \mathcal{R}^{3 \times 4}$ and the biases are $b_2 \in \mathcal{R}^4$, with the out puts being $O_2 \in \mathcal{R}^{p \times 4}$ and so on. The output layer will output $O_4 \in \mathcal{R}^{p \times 2}$, i.e. 2 output probabilities per input. Depending on which probability is higher for every example p_i, we get the prediction for that example.

3.2.2 Loss Functions

When a neural network is trained, a loss function, also known as cost function, needs to be defined to rate the efficiency of the network. This loss function can be of the form of L1 norms and L2 norms. These loss functions are used for regression problems such as predicting the price of a car given some input features. The L1 norm loss is the absolute value of the error in the predicted value, while the L2 norm loss (or the squared error loss function) is the square of the error in the predicted value. Let h_w be the prediction of the network. Mathematically these loss functions can be represented as:

$$\min \ \|h_w(x_i) - y_i\|_1$$

for L1 norm and

$$\min \ \|h_w(x_i) - y_i\|_2^2$$

for L2 norm.

The L1 norm is more robust than the L2 norm and gives much sparser outputs as compared to the L2 norm.

For classification problems, minimizing negative log-likelihood is used. Mathematically, we express it as:

$$\min \ -\log(p_k),$$

where p_k is the probability of the true label and $(1 - \sum_{i=1, i \neq k}^{n} p_i)$ is the probability of the incorrect predictions. Instead of using scores as predictions, a softmax function is applied for every score of an example to convert scores to probabilities. If $p_k = 1$ then the loss is minimized. Otherwise the lower the probability of p_k, the larger is $-\log(p_k)$, and thus higher losses.

3.2.3 Regularization

Loss functions help in training the model on the dataset such that the model can make the best predictions on the data it has been trained on. But our goal is to ensure that the model does well on all future data, most of which will not have been seen by the model. Regularization helps in putting constraints on the learning process of the model so it does not overfit to the dataset it is being trained on and learns general features for the prediction problem. Usual regularizers include L1 norm, L2 norm, and Dropout. L1 norm is usually used for sparse features while L2 norm is used to minimize magnitude of feature weights.

3.3 Training Neural Networks

3.3.1 Stochastic Gradient Descent

One of the primary methods of learning weights in machine learning is stochastic gradient descent. It involves finding the gradient of the expected loss of the current input and updating the weights in the direction that minimizes this loss. Mathematically, we represent it as:

$$w \leftarrow w - \alpha \frac{\mathcal{E}[\partial J(w)]}{\partial w}$$

where α is the learning rate and is a hyper-parameter. When a batch of inputs is passed in, we get:

$$w \leftarrow w - \alpha \frac{\mathcal{E}[\sum_{i=1}^{n} \partial J_i(w)]}{\partial w}$$

3.3.2 Backpropagation

When stochastic gradient descent is applied to neural networks, we need to find the gradients of the loss $J(w)$ with respect to the weights w of the respective layer. The naive method would be to recompute the gradients at each layer because it is computationally expensive and slow. Hinton et. al [18] then came out with their backpropagation algorithm, which made computing gradients quick. Using chain rule, one can compute the gradients for W_3 in Fig 3.2 like the following:

$$w \leftarrow w - \alpha \frac{\mathcal{E}[\partial J(W_4)]}{\partial W_4} \frac{W_4}{\partial W_3}$$

We know that gradient $\frac{\mathcal{E}[\partial J(W_4)]}{\partial W_4}$ is computed for the output layer already. Thus, by only computing the gradient of the next layer weights W_4 with respect to the previous layer W_3 and reusing the previously computed gradient, we can compute the overall gradient.

3.4 Convolutional Neural Networks

3.4.1 Importance

Convolutional neural networks are specialized neural networks used in situations where the input is an image or has important spatial information to be learnt. They can perform tasks such as image classification, object recognition and clustering of images based on similarity. These neural networks are very powerful and efficient and have a wide range of application, ranging from self-driving cars to detection of diseases in the field of medicine.

Compared to a normal neural network, convolutional neural networks are much more effective in feature extraction. This is not only because of their use of filters of various sizes to map out each channel of the image effectively but also because useful data is not lost when the image is flattened out into a vector before passing it to the classification layer. In normal neural networks, images are flattened out into vectors when they are received as inputs and this may cause a loss of useful spatial data which may adversely affect the accuracy of the neural network.

3.4.2 Defining the Model

There exist 3 important layers: Convolutional Layers, ReLU and Max-Pool. Using a combination of these layers are responsible for extracting features from images. After extracting these features, we attach a few layers of a normal neural network and then a classification layer.

A Convolutional Layer is categorized be a filter $w \in \mathcal{R}^{k\times k\times C}$ that is used on the image $I \in \mathcal{R}^{H\times W\times C}$. Note that the filter needs to have the same depth as the image, so as to average the features for all color pixels. Let us now take the filter on the top-left corner and find the dot product of the w with the region of the image that w overlaps with. This gives us one feature of the output of the layer. Then, depending on the stride, we shift the w to the right by those many pixels and then perform another dot product to get another output feature, and so on. When we reach the end of the image, we shift down by the appropriate number of pixels and continue the same process. Let the strides be s_1 and s_2 on each side respectively. Mathematically, the size of each side of the output layer will be:

$$d_1 = \frac{(H-k)}{s_1}$$
$$d_2 = \frac{(W-k)}{s_2}$$

The image below demonstrate the process visually:

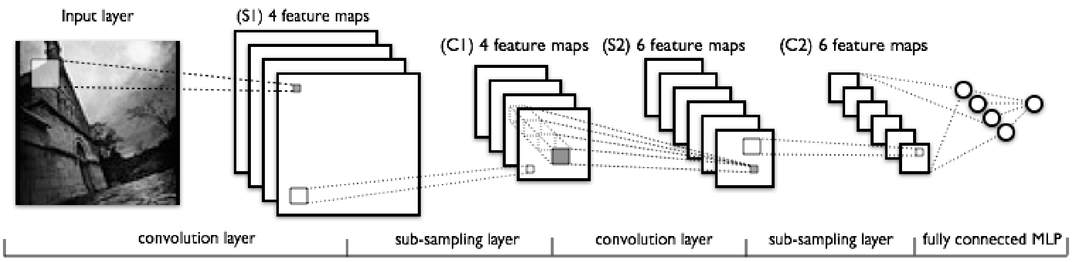


Figure 3.3: Working of Convolution Neural Network, [7]

If we had strides of $m \times n$, then we would skip m rows and n columns of pixels, before finding the next dot product. This makes the strides a hyper-parameter.

3.5 Performance Enhancing Techniques

3.5.1 Dropout

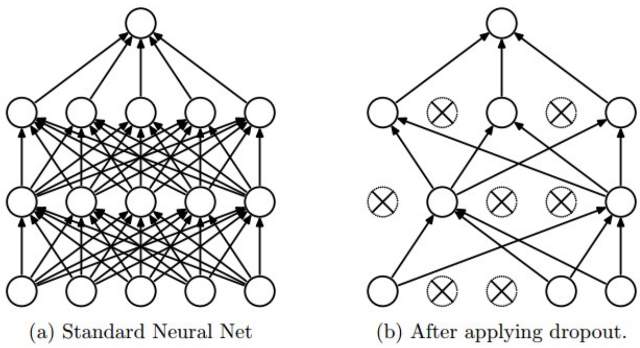


Figure 3.4: Dropout, from Department of Computer Science, University of Toronto [20]

Dropout is a regularization technique which is used to reduce over-fitting in neural networks. In this technique, nodes are randomly 'turned off' in a particular hidden layer with a probability p. By 'turning off', it is meant that the output of the nodes is nullified and does not affect the input to the next layer of the neural network. Dropout is implemented to ensure that all nodes in the neural network do not specialize in predicting the training set, as they are turned off randomly during the learning process. This ensures that over-fitting is reduced. That being said, it is important to keep in mind that dropout is not implemented while the neural network is being used for prediction as it may then adversely affect the accuracy of the predictions of the neural network.

3.5.2 Batch Normalization

Batch normalization is a technique which is used to reduce internal co-variate shifts during the training of neural networks. Internal co-variate shift refers to the change in the shape of the distribution as the inputs pass through different layers of the neural network. Due to this internal co-variate shift, the neural network does not learn properly as the user input to the network is different from the input passing through the network. Batch normalization is employed to reduce this shift and help the neural network converge faster.

In this technique, the normalization does not occur in the entire training set. Instead, the mean and variance of a given set of input or a mini-batch is computed, so that they can be normalized effectively.

Batch normalization can be mathematically represented as [15] :

$$\mu_\beta \leftarrow \frac{1}{m}\sum_{i=1}^{m} x_i \qquad \text{mini-batch mean}$$

$$\sigma_\beta^2 \leftarrow \frac{1}{m}\sum_{i=1}^{m} (x_i - \mu_b)^2 \qquad \text{mini-batch variance}$$

$$\widehat{x}_i \leftarrow \frac{x_i - \mu_\beta}{\sqrt{\sigma_\beta^2 + \epsilon}} \qquad \text{normalization}$$

$$y_i \leftarrow \gamma \widehat{x}_i + \beta \equiv BN_{\gamma,\beta}(x_i) \qquad \text{scale and shift}$$

3.6 Autoencoders

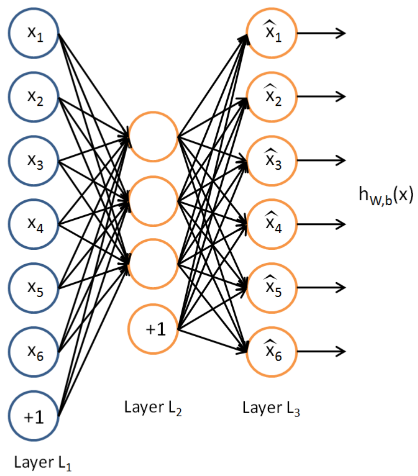


Figure 3.5: Autoencoder, from Deep Learning Tutorial, Stanford University [11]

An autoencoder is a symmetrical neural network which consists of 2 parts: the encoder and the decoder. The encoder consists of layers which reduce the dimensionality of input. The input is reduced to a vector in the case of normal neural networks, while in the case of a convolutional neural network the dimensionality of the input image is reduced.

In normal neural networks, the encoder reduces the number of hidden units in every successive layer to scale down the input. On the other hand, in convolutional neural networks, the encoder uses a filter with a strider greater than 1 to reduce the dimensionality of the input image. The decoder does the reverse of the encoder and attempts to convert the vector or image into the original input. There is an increase in the number of hidden units in every successive layer for a normal neural network, while there are deconvolutions along with a stride greater than 1 for convolutional neural networks. This serves to decode the reduced vector or images to the input matrix or image.

Reducing the dimensionality of the input helps capture important features and train neural networks faster. As a result, autoencoders play a major role in most implementations of convolutional neural networks.

Chapter 4

Autism Prediction using Neural Networks and Regression

4.1 Dataset and Pre-processing

The ABIDE data is a consortium of resting-state functional magnetic resonance imaging (R-fMRI) data sets from individuals with ASD collected by UCLA, Caltech and CMU over a period of few years. Multiple datasets are available on the ABIDE website, including phenotype and Region-of-Interest (ROI) data, but for our purposes, we only use the fMRI and the phenotype data.

4.1.1 Dataset

As described above, our dataset consists of fMRI data of patients with and without Autism. There is a 55% - 45% split between patients with autistic and control patients. The age of patients ranges from 10 months to 16 years old. Each fMRI image is a 91x108x91 3D image of the patients' brain and there are 1071 images in the dataset.

The phenotype data contained multiple attributes such as *DSM-IV TR* Diagnostic Critera, age, gender, handedness, FIQ (Full-scale IQ), VIQ (Verbal IQ) and PIQ (Performance IQ) standard score, and various Autism Diagnostic Interview scores.

4.1.2 Data Augmentation

There are only 1071 images in the dataset, and since neural networks need large amounts of data, data augmentation is essential. We want to use information about the nature of autism to drive our method to augment data. Based on our discussion in chapter 2, we know that ASD is a macroscopic neurological disorder, affecting functionality of brain functions. The dataset gives us the coordinates for the 116 regions of the brains, each region responsible for some brain process.

Now let us take 2 autistic brains. Our hypothesis is that autism, in general, affects similar regions of the brains, though it might be to different extents. We take corresponding regions of the 2 brains and randomly swap about 25% of the fMRI brain regions to create a new brain, which should also be autistic. Following this pattern, we can keep randomly swaping a different 25% brain regions to create new brain samples that should be autistic. Using this methdology, we use different combinations of the autistic and control brains to create 4815 fMRI images, that increases our training and test set.

4.1.3 Pre-processing

fMRI Images

After creating our augmented dataset, we want to be able to make the images ready to be able to train and test on. We normalize the images to prevent the networks from learning extreme pixels as patterns and to make the network be insensitive to base pixel values. Normalization is done using the following:

$$x_i^\star = \frac{x_i - \mu_i}{\sigma_i^2}$$

This means that we want mean 0 and a variance 1 images to learn identically scaled images.

Phenotype Data

The phenotype data contained several attributes, some of which had many incomplete values in various training examples. Due to this, the phenotype data had to be processed to keep only those attributes which had a good amount of data which could help in the diagnosis of autism. Finally, the phenotype data was processed to contain 497 total examples. The attributes were also now restricted to FIQ, VIQ, PIQ, age, gender, handedness.

4.2 Methodology

4.2.1 Convolutional Autoencoder

As mentioned in section 3.6, autoencoders are used to perform dimensionality reduction on inputs. Additionally, autoencoders preserve important features in the original data, hence information which may be important in predictions is not lost while encoding images using autoencoders. fMRI images are high dimensional images. Due to this, convolutional autoencoders are used to effectively encode and perform dimensionality reduction on the images.

The Convolutional autoencoders are based on autoencoders but have an important difference. Similar, to convolutional neural network, convolutional autoencoders (CAE's) make the use of convolutional filter to take advantage of the spatial structure in the input. Thus, as CAE's perform dimensionality reduction on data while at the same time preserving spatial relationship, the reduced data obtained from CAE's can be used as inputs to the Convolutional Neural Networks.

The cost function of the CAE is different from a convolutional neural network. A L2 norm between the original data and the encoded data is used. A KL-Divergence term is also added to the optimization objective in order to penalize the network for deviation of the ratio of the active hidden units from the sparsity parameter ρ. Mathematically this can be represented as [11]:

$$\widehat{p} = \frac{1}{m}\sum_{i=1}^{m}[a_j^{(2)}(x^{(i)})]$$

where $a_j^{(2)}$ denotes activation of hidden unit j in the autoencoder and $\widehat{p}$ is the average activation of each hidden unit j

The KL-Divergence term can be represented as:

$$\sum_{j=1}^{s_2} KL(\rho||\widehat{\rho}_j) = \rho \log \frac{\rho}{\widehat{\rho}_j} + (1-\rho)\log\frac{1-\rho}{1-\widehat{\rho}_j}$$

The overall cost function for the CAE can be represented as:

$$J_{sparse}(W,b) = J(W,b) + \beta\sum_{j=1}^{s_2} KL(\rho||\widehat{\rho}_j)$$

The final architecture designed and implemented is as follows:

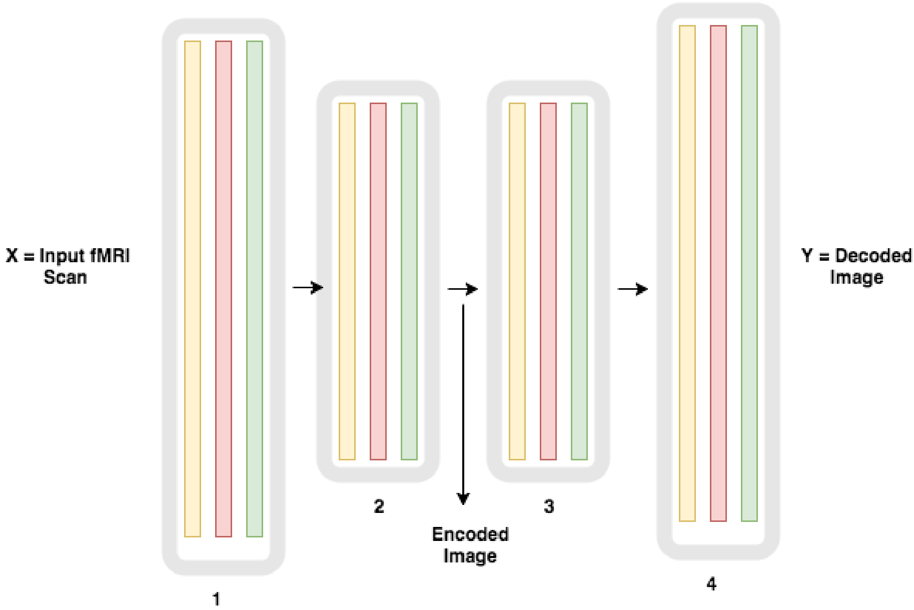


Figure 4.1: Overall Architecture of the AutoEncoder designed for best results and efficiency

1. Yellow is Convolutional Layer, having 3x3x3 filters with stride of 2.
   Red is Batch-Normalization layer.
   Green is ReLU Layer

2. Yellow is Convolutional Layer, having 3x3x3 filters with stride of 2.
   Red is Batch-Normalization layer.
   Green is ReLU Layer.

3. Yellow is Convolutional Layer, having 3x3x3 filters with stride of 2 for up-sampling (i.e. increasing size of the image) by factor of 2 (This is also called Deconvolution layer in the Keras package).
   Red is Batch-Normalization layer. Green is ReLU Layer.

4. Yellow is Convolutional Layer, having 3x3x3 filters with stride of 2 for up-sampling.
   Red is Batch-Normalization layer.
   Green is ReLU Layer.

A lot of Convolutional AutoEncoder architectures were tried. Below is an overview of our efforts:

- Having only 1 Convolutional, Batch Normalization and ReLU layer followed by 1 De-Convolutional, Batch Normalization and ReLU layer did not give good encoded images, even after training several times and for many days.

- Having only 3 Convolutional, Batch Normalization and ReLU layers followed by 3 De-Convolutional, Batch Normalization and ReLU layers did give good encoded images. But it took much longer to train and there was no difference between the images from the above architecture used and this deeper architecture. A smaller architecture is always preferable since it has less variables, captures the same information, is easier to train and uses less compute.

4.2.2 Convolutional Neural Network

A Convolution Neural Network (CNN) was used to learn spatial features across the image data. The CNN was used to take images from the Convolutional Autoencoder. Our architecture was as follows: each convolution unit had a micro-architecture of Convolution $\rightarrow$ Batch Normalization $\rightarrow$ ReLU.

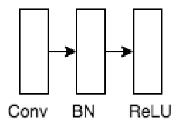


Figure 4.2: Micro-architecture of CNN : One Convolutional Unit, containing Convolution3D (Conv), Batch Normalization (BN) and ReLU (Rectified Linear Unit) layers

The convolution units were repeated 8 times in the macro-architecture of Convolution Unit $\rightarrow$ Convolution Unit $\rightarrow$ Max Pool. The first two convolution units had $5 \times 5 \times 5$ filter size while the rest had $3 \times 3 \times 3$ filter size.

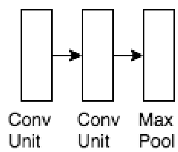


Figure 4.3: Macro - Architecture of CNN

Thus, we finally decided to use the following architecture:

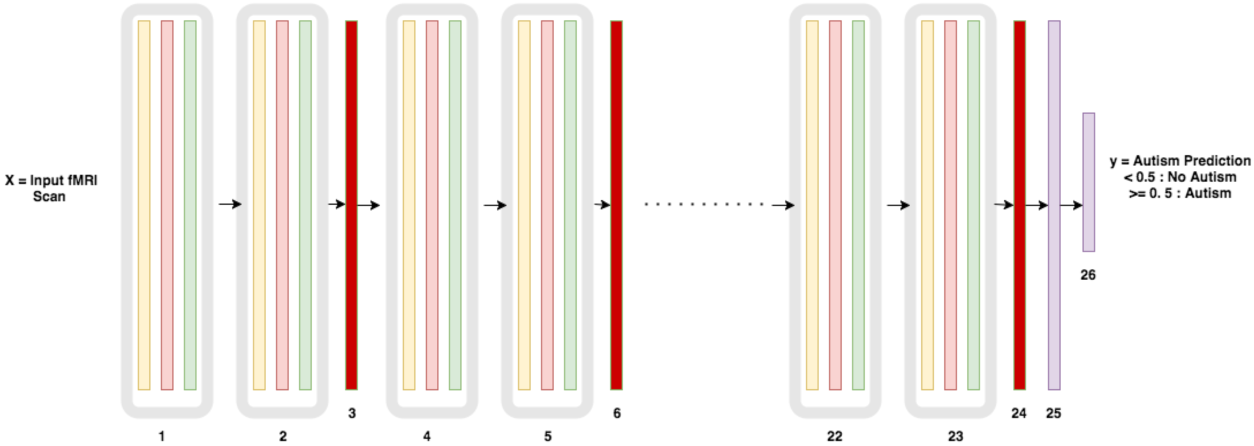


Figure 4.4: Overall Architecture of the Convolutional Neural Network

1 Yellow is Convolutional Layer, having 5x5x5 filters with stride of 1.
Light red is Batch-Normalization layer.
Green is ReLU Layer

2 Yellow is Convolutional Layer, having 5x5x5 filters with stride of 1.
Light red is Batch-Normalization layer.
Green is ReLU Layer.

3 Dark Red is Max-Pool layer of stride 1

4 Yellow is Convolutional Layer, having 3x3x3 filters with stride of 1.
Light red is Batch-Normalization layer.
Green is ReLU Layer

5 Yellow is Convolutional Layer, having 3x3x3 filters with stride of 1.
Light red is Batch-Normalization layer.
Green is ReLU Layer.

6 Dark Red is Max-Pool layer of stride 1

$\vdots$ (The above 3 blocks of layers repeated 6 more times)

$\vdots$

25 Fully Connected layer with 32 hidden neurons

26 Fully Connected layer with 1 neuron, having a sigmoid loss function. If the output is < 0.5, then the image is not an Autistic scan. If the output is ≥ 0.5, then the image is an Autistic scan.

Again, a lot of different Convolutional Neural Network architectures were tried:

- Instead of having 8 blocks of the Convolutional Unit, we tried to have blocks ranging from 2-7, but these architectures gave lower test-set accuracies that the one given by the above shown architecture.
- we also tried to have 10, 12 and 14 Convolutional Units in the network, but the increase in test-set accuracy was negligible ($< 0.3\%$). Hence we decided to go with the network which had lower computational cost to save on training time and also lower the number of variables for the most compact model.
- Instead of defining the macro-architecture as Convolution Unit $\rightarrow$ Convolution Unit $\rightarrow$ Max Pool, we tried defining the macro-architecture as Convolution Unit $\rightarrow$ Max Pool. In fact, this macro-architecture gave much worse results since there were not enough convolutions before the max-pool layers to extract relevant features before choosing the important features.
- On removing Batch-Normalization, the results of the model were much worse. This might be because of the regularizing effect of Batch-Normalization, coupled with the acceleration of training. Maybe if we had left the model to train for many more hours/days, a model without Batch-Normalization might have reached the same accuracies.

4.2.3 Regression Analysis

Linear Regression involves modeling a relationship between an independent variable and dependent variables. In this case the dependent variables are the features while the independent variable is the classification problem (autistic or non-autistic).

There are several forms of regression which include L2 and L1 regression. In this research, L2 regression is used. The squared error between the prediction and data label is minimized in this form of regression. Depending on the type of problem, the features input for regression may be converted to polynomial, logarithmic or exponential features to model a variety of functions. These modified features are then input to the linear regression model to gain effective predictions. Since linear regression involves linear algebraic calculations and not complex training procedure such as neural networks, it is less computationally expensive than the latter and runs faster. At times, regression yields than neural networks when less data is available. Consequently, we decided to use this model for prediction of autism using the phenotype data.

4.3 Programming Implementation

Figure 4.5: Software Pipeline of the Neural Network Implementation

4.3.1 Data Processing: fMRI Scans

The ABIDE Data was stored as RData, which is the file extension used to store data processed by the R programming language. RData cannot be easily processed in python, the language the models are being implemented in. Consequently, an R Script was written to save the data in the numpy format which can easily be used in python.

While building the models, it was evident that the numpy file size was too large and the program was taking time to process the data. Hence, the numpy files were compressed to HDF5 files which are smaller in file size, allowing faster processing. This was done by writing a python script in which the numpy data was loaded and then the h5py python package was used to save the data as HDF5 files. To ensure that the models train properly, all the data is normalized before it is used i.e. all the fMRI scan inputs have a mean of 0 and a variance of 1.

4.3.2 Convolutional AutoEncoder

The AutoEncoder to reduce the dimensionality of the fMRI brain scans was implemented in python using the Keras neural network package. For each layer of the autoencoder, the filter size and subsample size were declared after considering their possible effect on training. The Convolution3D and Deconvolution3D functions from Keras were used to implement the model here.

To ensure that dimensionality reduction was adequately taking place in the AutoEncoder, a custom activity regularizer had to be designed and implemented using TensorFlow package to be used in the AutoEncoder. This activity regular had an important role of reducing output size after each layer, and thus was essential for the correct functioning of the AutoEncoder.

The AutoEncoder model training was executed by a separate function we wrote to create the model. To ensure that the data regarding the training of the model is well documented, we also programmed this function to record and save a log file of the training which kept track of the loss after each epoch of training. The AutoEncoder had to be trained several times on the data before a correct reduced representation of the data was found.

To check if the dimensionality reduction was performed correctly, the original brain scan, the encoded image and the decoded image were all visualized using a function we wrote in python and compared. This was done by using the matplotlib package. After this, the reduced data was saved to train and test the convolutional neural network.

4.3.3 Convolution Neural Network (CNN)

The CNN to train and make predictions for autism diagnosis was implemented in python using the Keras neural network package. Similar to the AutoEncoder, for every layer of the CNN the number of filters, the filter size and the subsample size had to be declared. The Convolution3D, MaxPooling3D, Activation and BatchNormalization functions were used to form the various layers of the network.

To train the CNN model, a separate function was written to first create the CNN model. Similar to the AutoEncoder, the function was programmed to document and save the log files of training so that any errors or trends in the same could be analyzed to yield more accurate predictions.

After the CNN finished training on the train data, the model had to be tested to check its accuracy. This was done by designing and writing a test function in python which compared the predictions of the model and the true data values to find the overall test and train accuracy of this model.

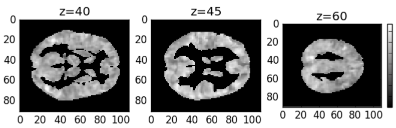
4.3.4 Regression

Before implementing the regression model, the phenotype data had to be processed. This was essential since a lot of the features in the dataset were incomplete and thus we manually processed the 1072 different examples in the CSV data to include only a few relevant features and examples for training and testing purposes. After this processing, the features in the processed dataset had to be encoded into a numpy array. Hence, we wrote a function to load the data and perform the encoding on the data before using it for regression analysis.

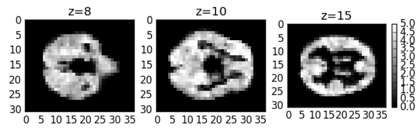
The regression model was programmed and implemented using the scikit-learn package. Several types of regression models were programmed and before being tested using a test function we wrote.

4.4 Results

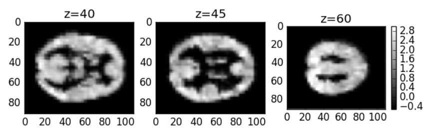
4.4.1 Convolutional AutoEncoder



(a) Input Image



(b) Encoded Image



(c) Decoded Image

Figure 4.6: Results of CAE

We see that the convolution auto-encoder was able to reduce the dimensionality of the fMRI 3-D scan from 91x108x91 to 31x37x31 clearly. The input image in many cases was also blurry, yet the output image was clear, as seen in Fig. 4.5.

4.4.2 Convolution Neural Network

After training the convolution neural network on the fMRI brain scans, the train and test accuracy were obtained.

Train Accuracy	Test Accuracy
0.9863	0.5486

Table 4.1: Accuracy of Convolution Neural Network

From table 4.1, we see that although the convolution neural network is able to learn the features for the train set which is signalled by its high train accuracy, it is not able to generalize what it has learnt, resulting in a low test accuracy. This also shows us the complexity of diagnosing autism and how brain structure can vary substantially for individuals.

4.4.3 Regression

Several types of regression such as polynomial and linear regression were performed on the processed phenotype data. This resulted in various diagnosis accuracies depending on the type of regression used.

Regression Type	Train Accuracy	Test Accuracy
Linear Regression	0.5568	0.6315
Polynomial Regression (Power = 2)	0.6045	0.6315
Polynomial Regression (Power = 3)	0.6386	0.6140
Polynomial Regression (Power = 4)	0.6772	0.7017
Polynomial Regression (Power = 12)	0.7	0.6842

Table 4.2: Accuracies of different types of Regression

From table 4.2, we can see that polynomial regression can be used to diagnose autism with an accurate diagnosis rate of around 70% , which is be higher than normal prevailing diagnosis rate for autism.

4.5 Analysis and Implications

4.5.1 Convolution Neural Network (CNN)

The CNN did not perform very well for the Brain fMRI scans. Looking back at the results and experiments tried, it could be as a result of the following:

1. There is a CNN architecture that has not been explored but might work well for this dataset. There are innumerable architectures that can be tried, and more study needs to be done as to what kind of architectures could be used and work best. Better models could be found out by doing a statistical analysis of how each neuron fired when an image was inputted, but such a study is beyond the scope of this project.

2. The dataset was not large enough, as a result of which the CNN could not learn relevant features and patterns to predict a complex condition like ASD. Maybe we need more augmentation and a larger supervised dataset.

The failure of the CNN model does not mean that a CNN cannot be used later. It just hints at the need for further and more detailed study, which we have managed to narrow down the search space to.

4.5.2 Regression

A regression analysis managed to give us results better than modern clinical success rates. Surprisingly, only simple and few features were needed to get these results. This means that maybe just a few well-filtered features are good enough indicators. This being said, it must be noted that autism may manifest itself in a similar way to other developmental disabilities in the phenotype data and hence the regression model needs to be tested on datasets of other disorders to ensure that it predominantly gives a unique prediction for autism.

On one hand, regression is less computationally expensive and easier to use than Deep Networks, but creating the required features requires manual human effort, which is tedious and might be inconsistent across different doctors. On the other hand, CNNs are more computationally expensive but since fMRI scans are standardized across the world and no features need to be created, the predictions are accurate and consistent.

Chapter 5

Case Study: India

5.1 A Brief Outlook

Autism Spectrum Disorder (ASD) affects around 10 million people [1] in India. In addition to the lack of infrastructure and trained medical personnel to effectively handle the the situation of ASD in India, the social stigma attached to autism and other developmental disorder in general worsens the situation not only for the patient but also for the family of the patient. ASD is one of the less known disorders in India but is also one which is very prevalent among the people. Large scale studies must be conducted in India to find out the problems in the ASD response system of the country, so as to improve the lives of the affected people.

5.2 Current Situation

5.2.1 Prevalence

According to the World Health Organization (WHO), on an average 1 in 160 [4] children worldwide have ASD. The exact prevalence rates of autism vary from region to region. India is one of the countries where epidemiological work on ASD needs to be conducted in order to get an accurate prevalence and incidence rate of autism. At the moment, no large scale study to find out the rates has been conducted in India. Although, estimates put the prevalence rate between 1 in 500 and 1 in 150. A study by the International Clinical Epidemiology Network Trust (INCLEN) found that about 1 to 1.5 % of children in India have autism. This would put the prevalence rate at about 1 in 66 [10], which is much higher than the global average. These statistics clearly show that ASD is a growing problem in India and must be tackled effectively through better diagnostic measures and treatment programmes.

5.2.2 Diagnosis

One of the main challenges which people affected by ASD face in India is getting an accurate diagnosis of the disorder. Many time, parents of ASD pateints are told that their children are 'slow', mentally retarded or schizophrenic [3] and due to this an accurate diagnosis of ASD is delayed. Consequently, valuable time lost which could have been used to start therapy for the patient and improve the prognosis of his disorder, is lost due to the delayed diagnosis.

A huge challenge which must be overcome to achieve the goal of accurate diagnosis of ASD in India is that more awareness about ASD and its diagnostic criteria must be there among medical professionals in the country. This will ensure an increase in the rate of accurate diagnosis in India. Apart from major cities and small pockets spread across the nation, most of rural India and small towns do not have professionals who can diagnose or deal with ASD cases [9]. Due to this, ASD is not usually diagnosed in rural India. This amplifies the need for an accurate diagnostic method in India which can help professionals effectively diagnose autism.

5.2.3 Treatment

In India, a few schools provide a wide range of special services to autistic children. These schools range from mainstream schools to autism-specific schools. This being said, the number of schools is inadequate relative to the number of autistic children in India [9]. Also, in most cases these schools are located in major towns and due to this autistic children in rural areas are severely affected by the lack of educational and therapy options in their region. Government schools many times are not equipped to handle autistic children and hence the interests of many autistic children in the country is not taken care of.

A few non-governmental organizations (NGOs) have started working in this sector, to help autistic children and other children with special needs. Although these organizations do a good job in helping the autistic children, there are not NGOs to effectively help autistic children in India. Also as ASD is a life-long disorder it is important that centres of care are established for autistic people where they can support throughout their life. Life can especially get tough for autistic people after their parents pass away and hence measures must be put into place to ensure that these people are given enough support so that they can reach their maximum potential.

5.2.4 Awareness and Acceptance

The lack of requisite awareness about autism among most most health professionals and people in India makes it really hard to diagnose ASD accurately in

India. Apart from this, it also makes it difficult for autistic people to be accepted by society as there traditionally has been a stigma attached to such disorders. To counter this social stigma, it is necessary that awareness campaigns about disabilities in general are held in educational institutions and places of work.

The awareness about autism is certainly greater than it was a decade or two earlier, but it seems that this has led to a new form of discrimination among educators where schools and other educational institutions sometimes turn away autistic students because they feel that they will not be able to provide individual support for them [9].

5.2.5 Legal Situation

ASD was recognized as a disability in India in 1999 through the 'National Trust for the Welfare of Persons with Autism, Cerebral Palsy, Mental Retardation and Multiple Disabilities Act, 1999,' which is more commonly known as the National Trust Act and then autism was officially recognized as a disability by the Government of India in 2001. Prior to that, autism was not recognized as a disability as it was not mentioned in the 'Persons with Disabilities (Equal Opportunities, Protection of Rights and Full Participation) Act' which was passed in the Indian Parliament in 1995.

There was no scale at that time to assess autism and due to this disability certificates were not being granted to autistic people. Finally after a 15 year wait and a period of intense lobbying, the National Institute of Mental Health came up with the Indian Scale for Assessment of Autism (ISAA) in 2009 for issuing disability certificates. However, the ISAA took 3 years to be notified before it was finalized in 2012 [5]. The wait of autistic people for disability certificates was not over yet. Finally in April 2016, The Department of Persons With Disabilities (DePwD) issued guidelines for issuing disability certificates to autistic people. Due to this, autistic people can now get disability certificates, which help them get several concessions such as exam accommodations, income tax concessions, among others [19].

5.3 Personal Experience

5.3.1 Ummeed Child Development Center

Ummeed is a non-profit organisation which helps children with developmental disabilities such as autism. I visited this center to learn more about ASD, its prevalence in India and to get a first hand perspective about it from the people who work at the center. I met with a member of the autism team at the center and was able to learn about specific details about ASD, while also understanding the grave situation regarding developmental disabilities in India.

During my conversation with the team member, I was able to gain a lot of insight into the current process of diagnosis of ASD and the emotional impact it has on the family of the diagnosed person. I was told that some parents go through a 'denial phase' when their children are diagnosed with ASD in India. I realized that this was due to the social stigma attached to such disorders and this emotionally touched me as I felt that due to such beliefs of society, autistic people and their families had to go through lot of social hardships.

I learnt a lot about the lack of facilities for autism therapy and diagnosis in Indian government hospitals. I also was able to gain insight into several severe problems regarding autism, such as families not turning up at autism centers after their child received a diagnosis of ASD and the lack of ASD diagnosis and therapy options in rural India, due to which probably the rising incidences of ASD cases is probably masked. I also got acquainted with the various autism therapy options offered, and how people go in for expensive treatments such as stem cell therapy without knowing that there is no proven effectiveness of the same, signaling the lack of awareness of autism treatment options and their effectiveness in India.

I also had a brief opportunity to witness a diagnosis session of autism and this helped me learn more about the system of diagnosis. On discussing the potential of my proposal to use fMRI scans to detect autism, I was told that it has potential and is a good thing to do in India where people usually 'forget' about people with disabilities.

I left Ummeed with a new outlook on the life of autistic people in India and their needs. I hope that through my research I will be able to help them by detecting autism early, so that early intervention can help improve their life.

5.4 Improving Diagnostic Methods

In order to improve the life of autistic people, it is important to diagnose ASD early. This early diagnosis, leads to an early intervention which has been proven to improve the prognosis of the disorder. Based on chapter 4, it can be said that polynomial regression on patient phenotype data have a huge potential in the early detection of ASD as results are quite accurate. On the other hand deep learning on fMRI scans was not able to diagnose autism accurately. Over here, there is scope for further research by training and testing the convolution neural network on fMRI scans of autistic and non-autistic people of a small age range such as 0-2,2-4, 4-6 years. This may lead to better results as a result of more similar brain structures.

Accurate and early diagnosis of ASD not only improves the life of autistic people due to early start of therapy but also reduce the overall costs of life-long support of these people by $\frac{2}{3}$. Though we did not get good results through the use of fMRI scans, they still have a lot of potential in autism diagnosis research, if sufficient data is available. This is not only due to the fact that fMRI scans are relatively safe as they use magnetic fields and not radiation to produce images, but also as fMRI scans can potentially reduce the time and effort needed to diagnose ASD and instead of a team of medical professionals diagnosing a person with ASD over multiple sessions, a diagnosis can be made accurately using neural networks on a scan. Thus use of fMRI scans and phenotype data in ASD detection should be further explored in research using more specific machine learning algorithms and a larger dataset while also experimenting with new techniques in the field.

Chapter 6

Conclusion

In this report, I have examined the autism spectrum disorder, neural networks and used regression and deep learning algorithms to try and detect autism accurately using fMRI scans and phenotype data respectively.

In case you have questions, comments, suggestions, please do not hesitate to contact me. You can find my contact details below.

Samir Agarwala
samir.aggarwala@gmail.com

Bibliography

[1] Santosh Andhale and Somita Pal. *Children with Autism can get better, shows Mumbai-based study*. 2015. URL: http://www.dnaindia.com/mumbai/report-children-with-autism-can-get-better-shows-mumbai-based-study-2073850.

[2] *Autism - Treatment Overview*. http://www.webmd.com/brain/autism/autism-treatment-overview. Accessed: 2016-12-22.

[3] *Autism is not a disease - Times of India*. URL: http://timesofindia.indiatimes.com/city/ahmedabad/Autism-is-not-a-disease/articleshow/19333810.cms.

[4] *Autism spectrum disorders*. URL: http://www.who.int/mediacentre/factsheets/autism-spectrum-disorders/en/.

[5] Merry Barua. *A disability certificate for autistic people will go a long way*. 2016. URL: http://www.hindustantimes.com/analysis/a-disability-certificate-for-autistic-people-will-go-a-long-way/story-wZe1f1HTudTknGmel97KXM.html.

[6] *Blood Test Shows Promise for Guiding Autism Diagnosis and Treatment*. https://www.autismspeaks.org/science/science-news/blood-test-shows-promise-guiding-autism-diagnosis-and-treatment. Accessed: 2016-12-22.

[7] *Convolutional Neural Networks (LeNet)*. URL: http://deeplearning.net/tutorial/lenet.html.

[8] *CS231n Stanford University*. http://cs231n.github.io/. Accessed: 2016-12-22.

[9] Tamara C. Daley. *Autism*. Ed. by Merry Barua. URL: http://www.rehabcouncil.nic.in/writereaddata/autism.pdf.

[10] Damayanti Datta. *Autism is rising alarmingly in India. How far is the new science from finding a cure? New research busts old myths and brings new hope*. 2013. URL: http://indiatoday.intoday.in/story/autism-autisitic-mind-western-syndrome-myths-about-autism-autistic-children/1/322242.html.

[11] *Deep Learning Stanford University*. http://deeplearning.stanford.edu/tutorial/. Accessed: 2016-12-23.

[12] *DSM-5 Diagnostic Criteria*. https://www.autismspeaks.org/what-autism/diagnosis/dsm-5-diagnostic-criteria. Accessed: 2016-12-22.

[13] *Facts and Statistics - Autism Society*. http://www.autism-society.org/what-is/facts-and-statistics/. Accessed: 2016-12-22.

[14] *How is Autism Treated*. https://www.autismspeaks.org/sites/default/files/documents/100-day-kit/treatment_version_2_0.pdf. Accessed: 2016-12-22.

[15] Sergey Ioffe and Christian Szegedy. "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift". In: *CoRR* abs/1502.03167 (2015). URL: http://arxiv.org/abs/1502.03167.

[16] Fuentes J et al. "Autism spectrum disorder". In: *Rey JM (ed), IACAPAP e-Textbook of Child and Adolescent Mental Health*. 2014.

[17] *Most Children With Autism Diagnosed at 5 or Older*. http://www.webmd.com/brain/autism/news/20120523/most-children-with-autism-diagnosed-at-5-or-older#1. Accessed: 2016-12-22.

[18] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. "Learning Representations By Back-Propagating Errors". In: *Nature* 323 (1986), pp. 533–536.

[19] Nidhi Sharma. *Autistic children to finally get disability certificates after a wait of 15 years*. 2016. URL: http://economictimes.indiatimes.com/news/politics-and-nation/autistic-children-to-finally-get-disability-certificates-after-a-wait-of-15-years/articleshow/52011831.cms.

[20] Nitish Srivastava et al. *Dropout: A Simple Way to Prevent Neural Networks from Overfitting*. Tech. rep. 1929-1958. 10 Kings College Road, Rm 3302, Toronto: Department of Computer Science, University of Toronto, 2014. URL: http://www.cs.toronto.edu/~rsalakhu/papers/srivastava14a.pdf.

[21] *The Importance of Early Detection*. http://www.parents.com/health/autism/symptoms/importance-of-early-detection-autism/. Accessed: 2016-12-22.

[22] *What the experts say: 4 new Autism research findings*. http://soniclearning.com.au/what-the-experts-say-4-new-autism-research-findings/, note = Accessed: 2016-12-22.