Evaluating the EEG and Eye Movements for Autism Spectrum Disorder

Sashi Thapaliya
Department of Computer Science
California State Polytechnic University
Pomona, CA 91768
sbthapaliya@cpp.edu

Sampath Jayarathna
Department of Computer Science
Old Dominion University
Norfolk, VA 23529
sampath@cs.odu.edu

Mark Jaime
Department of Psychology
Indiana Univeristy- Purdue University
Columbus, IN
jaime@iupuc.edu

Abstract— ***Autism Spectrum Disorder is a developmental disorder that often impairs a child's normal development of the brain. Early Diagnosis is crucial in the long term treatment of ASD, but this is challenging due to the lack of a proper objective measures. Subjective measures often take more time, resources, and have false positives or false negatives. There is a need for efficient objective measures that can help in diagnosing this disease early as possible with less effort. This paper presents EEG and Eye movement data for the diagnosis of ASD using machine learning algorithms. There are number of studies on classification of ASD using EEG or Eye tracking data. However, all of them simply use either Eye movements or EEG data for the classification. In our study we combine Eye movements and EEG data to develop an efficient methodology for diagnosis. This paper presents several models based on EEG, and eye movements for the diagnosis of ASD.***

Keywords—eye movements, EEG, autism spectrum disorder

I. Introduction

It is estimated that 1 in 6 children in the US suffer from developmental disorders. And 1 in 68 children fall under Autism Spectrum Disorder. Autism spectrum disorder is a neurological and developmental disorder that has negative impact in a person's learning, social interaction and communication. It is a debilitating condition that affects brain development from early childhood creating a lifelong challenge in normal functioning. Autism is measured in spectrum because of the wide range of symptoms and severity. One of the main contributing factor for ASD is known to be genetics. And so far no suitable cure has been found. However, early intervention has been shown to reverse or correct most of its symptoms[1]. And this can only be possible by early diagnosis. Therefore, early diagnosis is crucial for successful treatment of ASD. Although progress has been made to accurately diagnose ASD, it is far from ideal. It often requires various tests like behavioral assessments, observations from caretakers over a period of time to correctly determine the existence of Autism. Even with this tedious testing often times individuals are misdiagnosed. However, there remains promise in the development of accurate detection using various modalities of Biomedical Images, EEG, and Eye movements.

II. LITERATURE REVIEW

Studies have shown that EEG has the potential to be used as biomarker for various neurological conditions including ASD [2]. EEG measures the electrical signals of the brain via electrodes that are placed on various places on the scalp. These electrical signals are postsynaptic activity in the neocortex and can be used to study complex neuropsychiatric issues. EEG has various frequency bands and its analysis are performed on these varying bandwidths. Waves between 0.5 and 4 HZ are delta, between 4 and 8 HZ are theta, between 8 and 13 HZ are alpha, 13 to 35 HZ are beta and over 35 are gamma.

Saccadic eye movement plays a big role in the attention and behavior of an individual which directly affects both language and social skills [3]. Autistic children seem to have different eye movement behaviors than non-autistic children. They tend to avoid eye contact and looking at human face while focusing more on geometric shapes [4]. While a typical child doesn't find any interest in geometric shapes and tend to make more eye contact, and human face perception. We hypothesize that eye movements can be used as a biomarker for ASD which can be spotted early enough to be used as an objective measure for diagnosis.

In [5], authors use a complex EEG processing algorithm called MSROM/I-FAST along with multiple machine learning algorithms to classify Autistic patients. In this study 15 ASD individuals and 10 non ASD were selected. ASD group comprised of 13 males and 2 females between 7 and 14 years of age. Control group comprised of 4 males and 6 females between 7 and 12 years of age. Resting State EEG of both closed and open eyes were recorded using 19 electrodes. Patients sat in a quiet room without speaking or performing any mentally demanding activity while the EEG was being recorded. The proposed IFAST algorithm consists of exactly three different phases or parts. In the first stage also called Squashing phase, the raw EEG signals are converted into feature vectors. Authors present a workflow of the system from raw data to classification to make comparison between different algorithms such as Multi Scale Entropy (MSE) and the Multi Scale Ranked Organizing Maps (MS-ROM). MSROM is a novel algorithm based on Single Organizing Map Neural Network. In this study, the dataset is randomly divided into 17 training consisting of 11 ASD, 6 controls and eight test records consisting of 4 ASD, 4 control. The noise elimination is performed only on the training set. Also it completely depends on the algorithm selected for extraction of feature vectors. For MS-ROM features they utilize an algorithm called TWIST. In the final classification stage they use multiple machine learning algorithms along with multiple validation protocols. The validation protocols are training-testing and leave one out cross validation. For classification purposes they make use of Sine Net Neural Network, Logistic Regression,

978-1-5386-5035-6/18/$31.00 ©2018 IEEE

Sequential Minimal Optimization, kNN, K-Contractive Map, Naive Bayes, and Random forest. With MSE feature extraction the best results were given by Logistic and Naive Bayes with exactly 2 errors. Whereas, MS-ROM with training test protocol had 0 errors (100% accuracy) with all the classification models.

In [25], conduct a study using mMSE as feature vectors along with multiclass Support Vector Machine to differentiate developing and high risk infant groups. In this study they use 79 different infants of which 49 were considered high risk and 33 typically developing infants. The 49 infants were high risk based on one of their older siblings having a confirmed ASD diagnosis. The other 39 infants were not high risk based on the fact that no one in their family ever was diagnosed with ASD. Data was collected from each infant during multiple sessions with some interval. Data extracted from an infant in five different sessions in various months between 6 to 24 month period were considered unique. Resting state EEG with 64 electrodes was extracted by placing the infant in a dimly lit room in their mother's lap where the research assistant blew bubbles to catch their attention. The raw signals were preprocessed using Modified MultiScale Entropy. Low, high, and mean for each curve from mMSE were calculated to create a feature set of 192 values. The best fit for the classification for High risk and normal infants was at age 9 months with over 90% accuracy.

In [26], use EEG intrinsic function pulsation to identify patterns in Autism. They mathematically compute EEG features and compare ASD with typically developing. In this study they selected 10 children with ASD and 10 non autistic children within the age group of 4 to 13. They collected resting state EEG using 64 electrodes with a 500 HZ sampling frequency. Initially the signals were band pass filtered and all the artifacts including eye movements were removed by using Independent Component Analysis. Empirical Mode decomposition was applied to extract Intrinsic Mode Function from each of the channels of the participants. Then point by point pulsations of analytic intrinsic modes are computed which is then plotted to make comparison with the counterpart intrinsic mode in another channel. Any existing stability loops are analyzed for abnormal neural connectivity. In addition they perform 3D mapping to visualize and spot unusual brain activities. In the first IMF of channel 3 versus the first IMF in channel 2 for typically developing and autistic child it was found that the stability of local pulsation pathways maintained consistency while it was random in typically developing. Similar patterns were seen in channels 1 and 2 and 36 and 37 of non-autistic and autistic children. Overall this computational method was able to differentiate the abnormal EEG activities between ASD and typically developing children.

In [27] use Markov Models with eye tracking to classify Autism Spectrum Disorder. Unlike most other studies that collected data from children who were 3 years or older, in this study they collect data from 6 month old infants. There were in total 32 subjects out of which 6 were later at 3 years of age diagnosed with ASD and the rest were not. During the data collection the subjects were placed in front of their mothers and four different cameras from different angles recorded the video for about 3 minutes. The eye tracking was simply based on either the subject looked at the mother's face or not. Through this they get a binary sequence of subjects eye pattern which is then converted into alphabet sequence of a specific length. Then the sequence was filtered using a low pass filter and down sampled by factor of 18. This is done to enhance Markov Models to produce effective results. Using this data, they compare Hidden Markov Models and Variable-order Markov Models for the classification of ASD. Hidden Markov Models was able to correctly identified 92.03% of the typically developing subject while identifying only 33.33% of Autistic subject. Whereas the VMM correctly identified 100% of the Autistic and 92.03% of typically developing subjects. It was clear from this result that Variable-order markov models are superior in finding Autistic eye pattern while both Markov Models are the same in finding typically developing. The authors point out this difference as a result of various spectrums of Autism with different eye patterns. Nevertheless the VMM algorithm used in this study looks effective in identifying Autism in an early age.

Similarly in [28], propose a machine learning framework for the diagnosis of Autism using eye movement. They utilize two different datasets from previous studies. One of the dataset had 20 ASD children, 21 typically developing, and 20 typical developing IQ-matched children. The other dataset comprised of 19 ASD, 22 Intellectually disabled, and 28 typical young adults and adolescents. They compute Bag of Words for Eye Coordinates and Eye movement, N-Grams and AOI from the datasets. And they train five different Support Vector machine model with RBF kernel. Each of the model used different form of features like BOW of eye coordinates, BOW of eye movement, combination, N-Grams, and AOI. The end result was good for both groups with Combination or fusion data. However, the children dataset with fusion was the best with around 87% accuracy.

In another study [6], use eye movement with deep neural networks to identify individuals with Autism Spectrum Disorder. They used dataset from a previous study with 20 ASD and 19 health controls. Here the subjects observed around 700 images from the OSIE database. OSIE database is a popular eye tracking dataset used for image saliency benchmarking. First they use Cluster Fix algorithm on the raw data to compute fixations and saccades. Next, they work on finding the discriminative images as the OSIE dataset is not specifically built for autism studies. So, both groups might have the same visual pattern for some of the images. For this purpose they use Fisher score method by which they score each of the images and select only the one with the higher scores to be processed further. After this process of image selection they compute fixation maps in order to differentiate fixations between two groups. Fixation maps are simply a probability distribution of all the eye fixations. In addition they use a Gaussian Kernel for smoothing and normalize by their sum. Normalization is usually done when we are comparing two different fixation maps as is the case here. Then they compute difference of fixation map between the Autistic and non-Autistic group. This is the original target which they used to train a SALICON network to predict these values. SALICON network is one of the state of the art image

saliency prediction algorithm. Image saliency prediction is about predicting the visual pattern of users given an image. SALICON network uses two VGG with 16 layers. One of the VGG uses the original image to detect the small salient regions whereas the other VGG uses the down sampled image to detect the center of large salient regions. At the end both the outputs are combined to get a better result. This only predicts the image saliency. So in order to predict the difference of fixation map they add another convolution layer with Cross Entropy Loss function using the original Difference of fixation map. Next, they send the predicted difference of fixation maps to the final prediction layer. In this part they first apply tanh function to the features then concatenate the feature vectors of all fixation in order to consider dynamic change of attention. After which they reduce the dimension by using local average pooling. At last they train a SVM to make the final classification between ASD and control. They make use of the popular leave-one-out cross validation to measure the performance of their model. The accuracy of this model showed real promise in eye tracking for ASD with about 92% accuracy.

III. METHODOLOGY

Current techniques in practice for identifying ASD are mostly subjective and prone to error and usually takes a lot of time for final diagnosis. Most of the children with ASD are diagnosed after 3 years of age. Early diagnosis is the key for reversing or treating ASD through early intervention. As time is of an essence we need a method of diagnosis that is fast, and efficient unlike the current practice that could take months to years. Medical Imaging and blood testing [7, 8] are promising and a lot of work is being done with these modalities to diagnose ASD. However, EEG and Eye movement are cost effective and hence can be accessible in consumer level.

The aim of this research is to study the identification of Autism Spectrum Disorder using both EEG and Eye Tracking. Primary goal is to analyze ASD using EEG, Eye movement and combination of both. This will be done by classifying ASD using three different feature set, 1) only EEG, 2) only eye mmovement, 3) Combination of EEG and Eye Movement. Comparison of the classification performance between EEG, eye mmovement and combination of both EEG and Eye Movement can potentially result in finding the better feature set. We hypothesize as the top performing signal most likely has more of the unique data points and pattern of ASD and similarly, the least performing signals have less of the data points and patterns relating to ASD.

The secondary goal is to compare various machine learning algorithms like SVM, Deep Neural Network, Logistic, and Naive Bayes for the classification purposes. Conditions like ADHD, and other learning disabilities can also share similar comparative patterns for different features.

A. Dataset

We utilize the dataset from the work of [9] as this study collected both eye movement and EEG of both ASD and control group. Initially there are 52 participants consisting of 24 ASD and 28 control. Joint attention is the ability to socially coordinate visual attention, share a point of view with another person, and process self- and other-related information. Hence the data retrieval was performed while making the subjects watch video clips that would help in examining joint attention. There were a total of 12 videos each of which was 30 seconds. About one second gap was provided between each video. Both the EEG and Eye movement were collected while the participants watched the video. A total of 34 participants EEG data was used in this paper after the preprocessing.

A Tobi X50 eye tracker was used to collect the eye movement data. Visual fixation times were calculated using the eye trackers proprietary software: Clearview. The eye data comprises of stimulus data concatenated with the eye movement data. There were a total of 36 participants Eye data. The Social cognition measure test was also performed on the participants. This test helps in quantifying social cognition ability of the participants. Here, the subjects were each shown 28 pictures of the eye region and asked to describe the state of the person in one word. The more they correctly describe the actual state the higher the score. This feature was included in the eye movement data [9].

B. Preprocessing

For classification, we separated the data into about 80% training and 20% testing. 27 out of 34 were chosen for training while the rest 7 were used for testing. Amongst the 7 test data 4 were diagnosed with ASD while the other 3 were not. For this study, the original raw EEG data obtained in the study was preprocessed in EEGLab. Preprocessing of the raw EEG data consisted of both applying Makoto's Preprocessing Pipeline [10] and also using visual inspection to remove bad channels and data segments from the EEG data. The main purpose of this pipeline is to remove noise or artifacts, mainly eye blink from the EEG data.

There are many ways to extract feature from EEG data. Entropies, wavelets, FFT and various other statistical methods are commonly computed features. In this work we use Statistical and Entropy values. Statistical features comprises of Mean, Standard Deviation, and combined mean and standard deviation of the filtered data. Entropy is computed by using

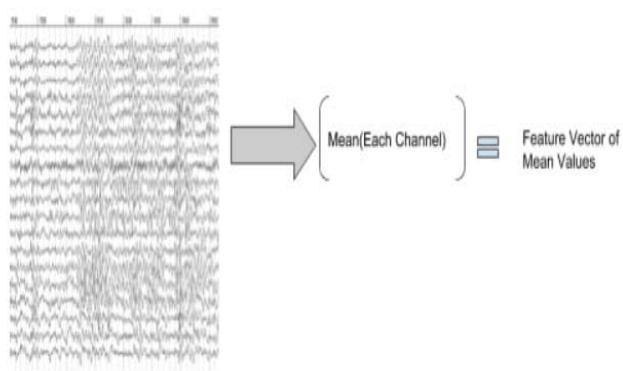

Fig 1. Raw EEG signal with 128 channels (2D matrix)-> Compute Mean of each of the channel and combine mean of all the channels vector with mean values of each channel

shannon entropy function [11] over windows of signal.

For the mean, each of the 128 channels were computed. For each subject a feature vector consisting of mean of single channel was created. So the mean function takes in a 2D matrix consisting of the EEG signal of a person and returns a feature vector with mean values for each channel. This is shown in the Fig1.

For the standard deviation, each of the 128 channels were computed. For each subject a feature vector consisting of mean of single channel was created. So the deviation function takes in a 2D matrix consisting of the EEG signal of a person and returns a feature vector with standard deviation values for each channel. This is shown in the figure below.

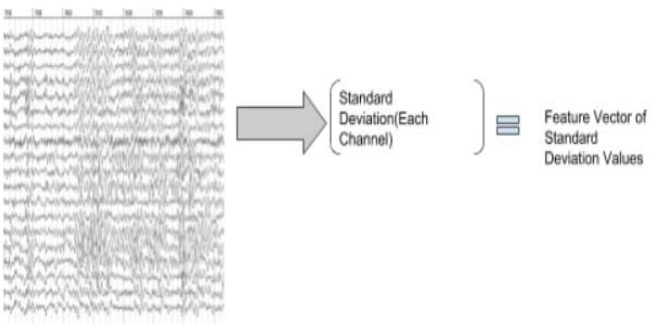

Fig 2. Raw EEG signal with 128 channels (2D matrix)-> Compute Standard Deviation of each of the channel and combine standard deviation of all the channels -> Vector with standard deviation values of each channel

Discrete Fast Fourier Transform [12] is computed using the formula below.

$$X_k = \sum_{n=0}^{N-1} x_n \cdot e^{-\frac{2\pi i}{N} kn}$$
$$= \sum_{n=0}^{N-1} x_n \cdot [\cos(2\pi kn/N) - i \cdot \sin(2\pi kn/N)],$$

$$M = \begin{bmatrix} FFTWindow1Channel1 & FFTWindow1Channel2 & \dots \\ FFTWindow2Channel1 & FFTWindow2Channel2 & \dots \\ FFTWindow3Channel1 & FFTWindow3Channel2 & \dots \\ FFTWindow4Channel1 & FFTWindow3Channel2 & \end{bmatrix}$$

Fig 3. Represents the FFT matrix of one subject. Each column represents a channel and rows represent a window.

The preprocessed signals are divided into windows of size 40 and step size 20. Then the DFFT is computed for each of the window for each channel. After this standard deviation and mean of the DFFT values of each of the channel is calculated. At the end we get a feature vector for each subject consisting of standard deviation and mean of DFFT values of each channel.

$$\text{Channel} = \begin{bmatrix} FFTWindow1 \\ FFTWindow2 \\ FFTWindow3 \\ FFTWindow4 \\ \dots \end{bmatrix} \Rightarrow \begin{bmatrix} SDFFT \\ MeanFFT \end{bmatrix}$$

Fig 4. Standard deviation and mean is computed for each channel separately

$$\begin{bmatrix} SDFFTChannel1 \\ MeanFFTChannel1 \\ SDFFTChannel2 \\ MeanFFTChannel2 \\ \dots \end{bmatrix}$$

Fig 5. Each feature vector consists of the mean and standard deviation of the windows of each channel.

Shannon Entropy "is the average rate at which information is produced by a stochastic source of data", it is computed using the formula below,

$$H = -\sum p(x) \log p(x)$$

The preprocessed signals are divided into windows of size 40 and step size 20. Then the entropy is computed for each of the window for each channel similar to DFFT.

For Eye movement the data points are fixation times along with the social cognition scores, age, and gender. No additional computation was performed.

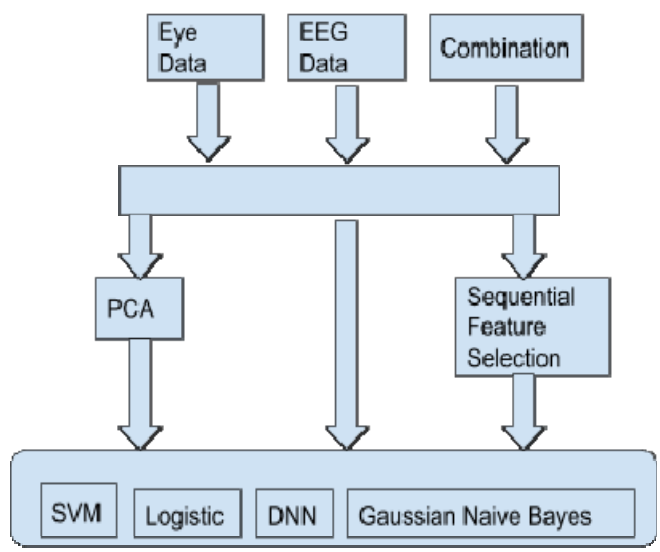

Fig 6. Methodology of classification

C. Models

Due to the 3 different types of datasets and the different mathematical features that we can extract from an EEG, there are many different types of data models (see Fig. 6).

Below is a high level overview of the classification pipeline using EEG.

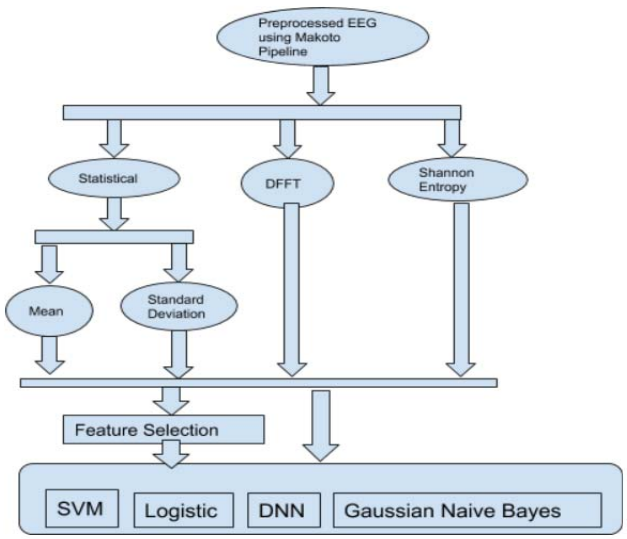

Fig 7. EEG classification pipeline.

As shown in the Fig 6. and 7., there are mainly three different feature set, entropy features, FFT and statistical features. In total there are 4 different features from EEG and 4 different models for each type of classifier. For classification SVM, Logistic, Deep Neural Network, and Gaussian Naive Bayes is used. Overall there are 16 different models. For each feature there are three models for each algorithm. Two models using Feature Selection and the third one without using any feature selection. For Feature selection PCA or Sequential Feature Selection method is used. All together there are more than hundred models.

Deep neural network with five hidden layers with sigmoid activation function was used. For optimization categorical cross entropy for loss and Adamax optimizer[13] was used.

D. *Classification using Eye Movement*

Fig. 8 is the high level overview of the classification pipeline. As shown in the figure, there are 4 classifiers working on the eye data. Eye data comprises of the Eye movement fixation times and the Eye test data.

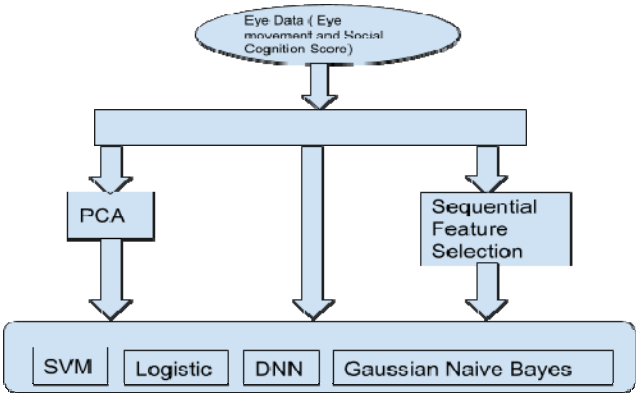

Fig 8. Classification Pipeline using Eye data

E. *Classification using Combination of EEG and Eye Data*

Fig. 9 is a high level overview of the classification pipeline using the combination of the EEG and the Eye data.

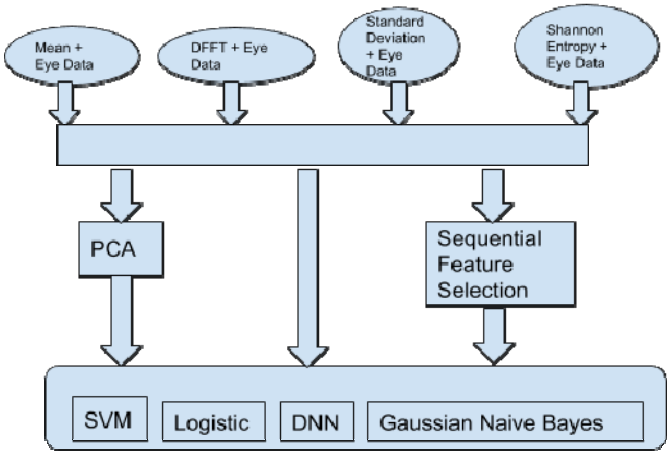

Fig 9. Classification Pipeline using combination of EEG and Eye Data

IV. RESULTS

Here is the comparison of all the different models. The best performance with the given algorithm and feature set is used in these tables below.

The first set of testing is done by classifying the testing data set. This section will show the various results while predicting on testing data.

The following table 1 is without using Sequential Feature Selection algorithm.

Table 1: Without Using Sequential Feature Selection Algorithm

	Only Eye Data	Only EEG SD	Combination
SVM	57%	57%	42%
Logistic	85%	71%	100%
DNN	71%	85%	85%
Gaussian Naive Bayes	100%	71%	71%

The above table compares between EEG standard deviation, Eye movement data and combination data models. SVM performed in the same level for both only Eye and EEG SD data. Logistic regression perform the best with 100% accuracy with combination data. DNN had same accuracy of 85% with both EEG SD and combination. Gaussian Naive Bayes had 100% accuracy with only Eye data. In this case, Gaussian Naive Bayes with only Eye data and Logistic regression with combined data were the better models with perfect classification. And it seems like both the Combination data and only Eye data perform about the same with good results.

Table 2: between EEG mean, Eye and combination data models

	Only Eye Data	Only EEG Mean	Combination
SVM	57%	42%	42%
Logistic	85%	57%	71%
DNN	71%	57%	42%
Gaussian Naive	100%	57%	100%

The above table 2 compares between EEG mean, Eye and combination data models. SVM with only eye data performed best with only 57% accuracy. Logistic Regression with only eye data performed best with about 85% accuracy. DNN with only eye data performed best with 71% accuracy. Gaussian Naive Bayes with only Eye data performed best with 100% accuracy. In this case with Eye data clearly is producing better results than the others.

Table 3: EEG FFT, Eye and combination data models

	Only Eye Data	Only EEG FFT	Combination
SVM	57%	42%	42%
Logistic	85%	71%	85%
DNN	71%	57%	71%
Gaussian Naive Bayes	100%	57%	71%

The above table 3 compares EEG FFT, Eye and combination data models. SVM with only Eye data do better than combined with the same accuracy. Logistic regression with combination and only Eye data have the same accuracy and beat only EEG FFT. Gaussian Naive Bayes with only Eye data performs the best with 100% accuracy. So, in this case Gaussian Naive Bayes Eye data is the winner.

Table 4: EEN Entropy, Eye and Combination of data models

	Only Eye Data	Only EEG Entropy	Combination
SVM	57%	71%	42%
Logistic	85%	71%	85%
DNN	71%	57%	71%
Gaussian Naive Bayes	100%	57%	100%

The above table compares EEN Entropy, Eye and Combination data models. SVM with only EEG Entropy has the best performance with 71% accuracy. Logistic Regression with only Eye data and Combination data have the same accuracy and beat only EEG Entropy model. Similarly, DNN with only Eye data and Combination data models beat only EEG Entropy data. Again Gaussian Naive Bayes with only Eye data and Combination data models have 100% accuracy and beat only EEG entropy model. In this case both the Eye data and combination data models seems to perform about the same and better than EEG entropy model.

Using Feature Selection Algorithm accuracy for both eye data and combination increases. The following data compare the results using Sequential Feature Selection algorithm.

Table 5: Eye data with Entropy EEG and combination

	Only Eye Data	Only EEG Entropy	Combination
SVM	76%	42%	100%
Logistic	100%	42%	100%
DNN	100%	57%	57%
Gaussian Naive Bayes	96%	28%	100%

The above table compares Eye data with Entropy EEG and combination. We can see that overall all algorithms except DNN have 100% accuracy. While Only Eye Data has Logistic and DNN with 100% accuracy.

Table 6: Eye data with FFT EEG and combination

	Only Eye Data	Only EEG FFT	Combination
SVM	76%	42%	42%
Logistic	100%	42%	85%
DNN	100%	71%	57%
Gaussian Naive Bayes	96%	42%	100%

The above table compares Eye data with FFT EEG and combination. We can see that only Eye data produces 100% accuracy for both Logistic and DNN. While Combination has Naive Bayes with 100% accuracy.

Table 7: Eye data with Standard Deviation EEG and combination

	Only Eye Data	Only EEG Standard Deviation	Combination
SVM	76%	42%	71%
Logistic	100%	57%	100%
DNN	100%	42%	57%
Gaussian Naive Bayes	96%	42%	100%

The above table compares Eye data with Standard Deviation EEG and combination. We can see that only Eye data produces 100% accuracy for both Logistic and DNN. Similarly, Combination has both Naive Bayes and Logistic with 100% accuracy.

Table 8: Eye data with Mean EEG and combination

	Only Eye Data	Only EEG Mean	Combination
SVM	76%	42%	42%
Logistic	100%	57%	100%
DNN	100%	57%	42%
Gaussian Naive Bayes	96%	57%	100%

The above table compares Eye data with Mean EEG and combination. We can see that only Eye data produces 100% accuracy for both Logistic and DNN. Similarly, Combination has both Naive Bayes and Logistic with 100% accuracy.

Average accuracy is shown in the table 9 and 10. All the models except DNN was run 200 times. DNN was ran only for 10 iterations due to the computationally exhaustive nature.

Table 9: 10x2 Cross Validation

Models	Eye	Entropy EEG	FFT EEG	SD EEG	Mean EEG	Combined
Gaussian Naive Bayes	96%	26%	53%	55%	55%	100%
Logistic Regression	100%	11%	78%	50%	58%	100%
SVM	76%	11%	56%	55%	55%	90%
DNN	100%	20%	52%	45%	58%	46%

Table 10: comparing the average of 10x2 Cross Validation for all the datasets

Models	Combined Eye + Entropy EEG	Combined Eye + FFT EEG	Combined Eye + Mean EEG	Combined Eye + Standard Deviation EEG
Gaussian Naive Bayes	100%	100%	80%	88%
Logistic Regression	90%	90%	83%	100%
SVM	90%	88%	40%	86%
DNN	43%	40%	46%	13%

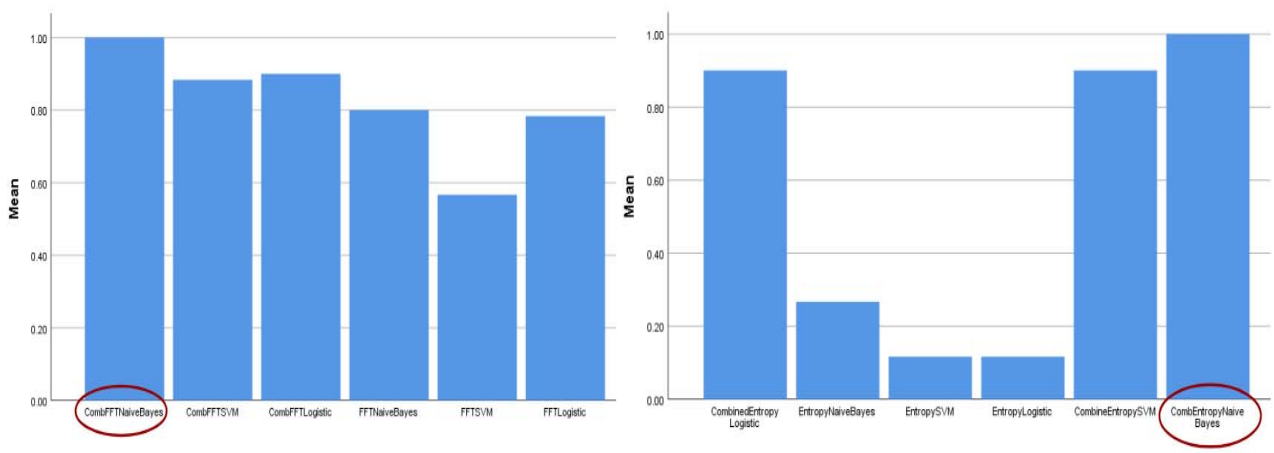

Fig. 10: (a) FFT EEG Vs Combined FFT EEG, (b) Entropy EEG Vs Combined Entropy EEG

In the above table 9 we are comparing the average of 10x2 Cross Validation that was run 200 times for all the varying datasets. We can see that combined produces 100% accuracy with Naive Bayes and Logistic Regression. Similarly, Eye produces 100% accuracy with Logistic and DNN.

Similarly in table 10 we are comparing the average of 10x2 Cross Validation for all the different combination data..

We can see that combined eye with Entropy and FFT EEG produces 100% accuracy with Naive Bayes. And Combined with Standard Deviation EEG produces 100% accuracy with Logistic regression. While Combined Eye with Mean has the highest accuracy of only 83% with Logistic Regression.

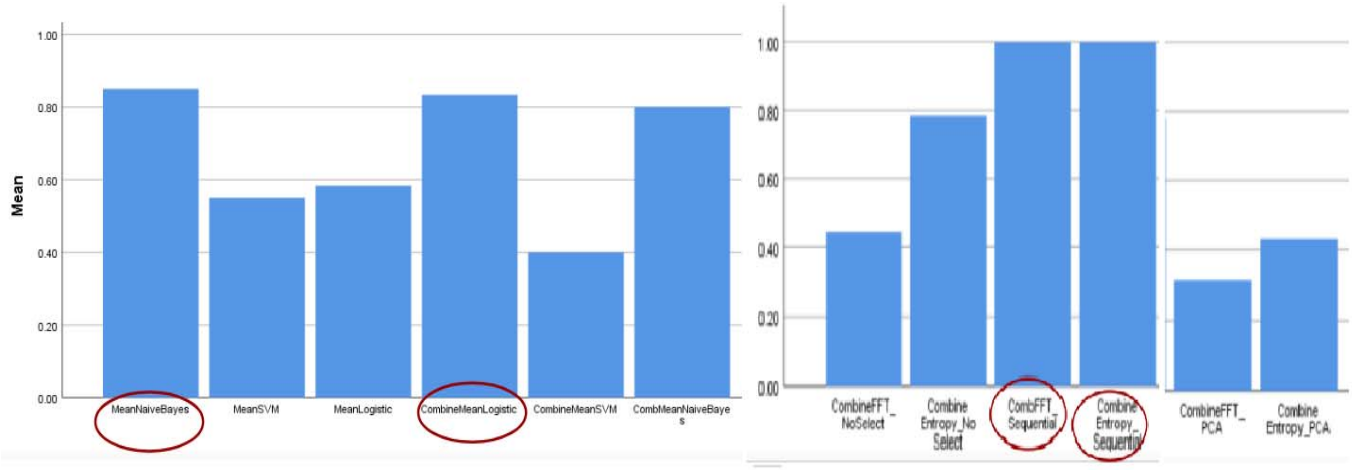

Fig. 11: (a) Mean EEG Vs Combined Mean EEG, (b) Feature Selection Algorithm Comparison

Figure 10 and 11 shows the comparison bar graph of FFT EEG and combined FFT EEG using Sequential Feature Selection.

In the Figure 10(a) we can clearly see than Combine FFT and eye data with Naive Bayes is performing highest with 100% average accuracy. And all the other Combined models are performing higher than the single EEG FFT models. In the Figure 10(b) that shows the comparison between Entropy EEG models and Combined Eye and Entropy EEG models. In that we can see than Combine Entropy Naive Bayes has the highest average accuracy of 100%. And all the combined models perform better than models with just Entropy EEG data. In the Figure 11(a) we can see than highest performing models with around 82% average accuracy are Naive Bayes with just Mean EEG and Logistic with combined data. In the Figure 11(b) we can see than Sequential Feature Selection algorithm performs the best.

V. CONCLUSION AND FUTURE WORK

This paper presents an analysis and comparison between EEG, Eye and combined data. We have compared four models with only Eye data, thirty two models with EEG data, and thirty two models with combined EEG and Eye data. Note that two models were created for each model with only EEG and combined data by using PCA and without using PCA. Like SVM with PCA and without PCA. For some models with PCA did better while for some without PCA did better. For example, DNN almost always without using PCA did worse because of the curse of dimensionality. The highest performing SVM with about 71% accuracy was using Shannon Entropy with all the features without PCA. The highest performing Logistic regression with 100% accuracy was using combination of EEG Standard Deviation and eye data without PCA. SVM, Logistic Regression, and Gaussian Naive Bayes do better without PCA which means that with PCA it loses data points that these models find useful. This is interesting because PCA is supposed to find the most discriminant features and remove redundant or noisy features. And this is supposed to help machine learning models produce better results. For SVM most models with PCA did better except the highest performing model. This might mean that the Entropy data is more linear than the other datasets. For DNN the curse of dimensionality is obvious. Whereas for Gaussian Naive Bayes all the high performing models did not use PCA except the one with the combination of EEG mean and eye data. This is an exception and must be due to the nature of the EEG mean data. But in general case Naive Bayes does better without PCA. This might be due to the fact that probabilistic models are able to make sense of higher dimensional dataset much easier than other models like DNN.

Then with using Sequential Feature Selection algorithm almost all the models performed better than either PCA or no Feature Selection. However, EEG data was still performing poorly than either Eye or combined data. In the future there are a lots of areas of improvement and a lot more comparison that can be made. In this study we have used PCA, and Sequential Feature Selection algorithm. There are other Feature Selection algorithms like Genetic algorithm, Particle Swarm Optimization, and TWIST which can be compared to find features to optimize the performance of the models. Also, this will tell us which feature selection algorithm will work better for the combined data sets. Gaussian Naive Bayes with some of the features had perfect score. But we need to reproduce this result with large number of participants to be able to use this in a clinical setting. Current number of 34 participants is too low to confirm our results. However, this is a first step towards developing an optimal Autism Diagnosis system.

REFERENCES

[1] Dawson, G., *Early behavioral intervention, brain plasticity, and the prevention of autism spectrum disorder.* Development and psychopathology, 2008. **20**(3): p. 775-803.

[2] Wang, J., et al., *Resting state EEG abnormalities in autism spectrum disorders.* Journal of neurodevelopmental disorders, 2013. **5**(1): p. 24.

[3] Fletcher-Watson, S., et al., *Eye-movements reveal attention to social information in autism spectrum disorder.* Neuropsychologia, 2009. **47**(1): p. 248-257.

[4] Klin, A., et al., *Two-year-olds with autism orient to non-social contingencies rather than biological motion.* Nature, 2009. **459**(7244): p. 257.

[5] Grossi, E., C. Olivieri, and M. Buscema, *Diagnosis of autism through EEG processed by advanced computational algorithms: A pilot study.* Computer methods and programs in biomedicine, 2017. **142**: p. 73-79.

[6] Jiang, M. and Q. Zhao. *Learning Visual Attention to Identify People with Autism Spectrum Disorder*. in *Proceedings of the IEEE International Conference on Computer Vision*. 2017.

[7] Sparks, B., et al., *Brain structural abnormalities in young children with autism spectrum disorder.* Neurology, 2002. **59**(2): p. 184-192.

[8] Spence, S.J., P. Sharifi, and M. Wiznitzer. *Autism spectrum disorder: screening, diagnosis, and medical evaluation*. in *Seminars in Pediatric Neurology*. 2004. Elsevier.

[9] Jaime, M., et al., *Brief report: reduced temporal-central EEG alpha coherence during joint attention perception in adolescents with autism spectrum disorder.* Journal of autism and developmental disorders, 2016. **46**(4): p. 1477-1489.

[10] Mullen, T.R., et al., *Real-time neuroimaging and cognitive monitoring using wearable dry EEG.* IEEE Transactions on Biomedical Engineering, 2015. **62**(11): p. 2553-2567.

[11] Lin, J., *Divergence measures based on the Shannon entropy.* IEEE Transactions on Information theory, 1991. **37**(1): p. 145-151.

[12] Weinstein, S. and P. Ebert, *Data transmission by frequency-division multiplexing using the discrete Fourier transform.* IEEE transactions on Communication Technology, 1971. **19**(5): p. 628-634.

[13] Freivalds, K. and R. Liepins, *Improving the Neural GPU Architecture for Algorithm Learning.* arXiv preprint arXiv:1702.08727, 2017.

[14]] 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. *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*. doi:10.1109/acii.2015.7344638

[15] Jiang, M., & Zhao, Q. (2017). Learning Visual Attention to Identify People with Autism Spectrum Disorder. *2017 IEEE International Conference on Computer Vision (ICCV)*. doi:10.1109/iccv.2017.354

[16] Wang, J., Barstein, J., Ethridge, L. E., Mosconi, M. W., Takarae, Y., & Sweeney, J. A. (2013). Resting state EEG abnormalities in autism spectrum disorders. *US National Library of Medicine*. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3847481/.

[17] Cantor, D. S., Thatcher, R. W., Hrybyk, M., & Kaye, H. (1986). Computerized EEG analyses of autistic children. *Journal of Autism and Developmental Disorders,16*(2), 169-187. doi:10.1007/bf01531728

[18] Training Attention and Eye Movement in Autism Spectrum Disorder. (n.d.). Retrieved December 27, 2017, from https://neurosciences.ucsd.edu/research/labs/autism-and-development/research/Pages/research_attentrain.aspx

[19] Jaime, M., McMahon, C. M., Davidson, B. C., Newell, L. C., Mundy, P. C., & Henderson, H. A. (2016). Reduced Temporal-Central EEG Alpha Coherence during Joint Attention Perception in Adolescents with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *46*(4), 1477–1489. http://doi.org/10.1007/s10803-015-2667-3

[20] P. Hoole *et al*., "Autism, EEG and brain electromagnetics research," *2012 IEEE-EMBS Conference on Biomedical Engineering and Sciences*, Langkawi, 2012, pp. 541-543. doi: 10.1109/IECBES.2012.6498036

[21] A. Sheikhani, H. Behnam, M. R. Mohammadi, M. Noroozian and P. Golabi, "Analysis of quantitative Electroencephalogram background activity in Autism disease patients with Lempel-Ziv complexity and Short Time Fourier Transform measure," *2007 4th IEEE/EMBS International Summer School and Symposium on Medical Devices and Biosensors*, Cambridge, 2007, pp. 111-114. doi: 10.1109/ISSMDBS.2007.4338305

[22] https://sccn.ucsd.edu/wiki/Makoto's_preprocessing_pipeline

[23] https://en.wikipedia.org/wiki/Entropy_(information_theory)

[24] Baron-Cohen S, Wheelwright S, Hill J, Raste Y, Plumb I. The “Reading the Mind in the Eyes” Test

[25] Bosl, W., Tierney, A., Tager-Flusberg, H., & Nelson, C. (2011). EEG complexity as a biomarker for autism spectrum disorder risk. *BMC Medicine,9*(1). doi:10.1186/1741-7015-9-18

[26] E. Abdulhay, M. Alafeef, H. Hadoush, N. Alomari and M. Bashayreh, "Frequency 3D mapping and inter-channel stability of EEG intrinsic function pulsation: Indicators towards autism spectrum diagnosis," *2017 10th Jordanian International Electrical and Electronics Engineering Conference (JIEEEC)*, Amman, 2017, pp. 1-6.

[27] D. Alie, M. H. Mahoor, W. I. Mattson, D. R. Anderson and D. S. Messinger, "Analysis of eye gaze pattern of infants at risk of autism spectrum disorder using Markov models," *2011 IEEE Workshop on Applications of Computer Vision (WACV)*, Kona, HI, 2011, pp. 282-287.

[28] 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. *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*. doi:10.1109/acii.2015.7344638