Evaluation of EEG and Eye movement with machine learning for the classification of Autism Spectrum Disorder

A Thesis

Presented to the Faculty of

California State Polytechnic University, Pomona

In partial Fulfillment

Of the Requirements for the Degree

Masters of Science

In

Computer Science

By

Sashi Thapaliya

2018

SIGNATURE PAGE

THESIS: EVALUATION OF EEG AND EYE MOVEMENT WITH MACHINE LEARNING FOR THE CLASSIFICATION OF AUTISM SPECTRUM DISORDER

AUTHOR: SASHI B THAPALIYA

TERM SUBMITTED: Summer 2018

Computer Science Department

Dr. Sampath Jayarthna
Thesis Committee Chair
Department of Computer Science

Dr. Gilbert S. Young
Department of Computer Science

Dr. Hao Ji
Department of Computer Science

ACKNOWLEDGEMENTS

First of all I would like to thank God for making be able to do this. I am so grateful to my advisor Dr. Sampath for providing all the guidance and support. I have learned a ton from Dr. Sampath about research, career in Computer Science and life in general. And I greatly appreciate Dr. Jaime's collaboration for this research. He provided the valuable data without which this work would not have been possible. And big thanks to my committee members Dr. Young and Dr. Ji for being awesome. Dr. Young was helpful in many ways as a committee member and Graduate advisor. And Dr. Ji taught me to conduct research in machine learning. And I would like to thank my family for all the support they provided while pursuing this MS degree. In addition I would like to thank my mentor Dr. Kazunori Okada for providing all the needed support and encouragement.

Overall, I am super grateful for all the support I had while pursuing my MS degree from everyone around me.

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 treatment of ASD. But currently this is very challenging due to the lack of a proper objective measure. Subjective measures often take more time, resources, and have false positives or false negatives. There is a need for objective measures that can help in diagnosing this disease early as possible with less effort. In this thesis I analyse EEG, and Eye movement data for the diagnosis of ASD using machine learning algorithms. There are number of studies that worked on classification of ASD using EEG or Eye tracking data. However, all of them simply use just Eye or EEG data only for the classification. None of them have used both Eye and EEG data at the same time. Yet, they are all promising and bring us to an important question if which set of data are better or worse for ASD classification. This is important to know in order to develop an optimal clinical system for diagnosis. Hence, in this thesis I analyse and compare EEG, Eye and combined EEG with Eye data for the diagnosis of ASD. I perform this comparison by using dataset collected by Dr. Jaime and colleagues in another study where both Eye and EEG data of adolescent participants were collected at the same time. For classification I use algorithms like SVM, Logistic Regression, Deep Neural Network, and Gaussian Naive Bayes.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

According to CDC 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. And this can only be possible by early diagnosis. So, 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 movement. In this thesis I would like to make an analysis of using EEG and Eye tracking in identifying ASD.

1.1 EEG

Studies have shown that EEG has the potential to be used as biomarker for various neurological conditions including ASD.[16] 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.

1.2 Eye Movement

Eye movement has shown to contain biometric features for an individual. Studies has shown that it is a good indication of neurological disorders as well. In fact it is used by medical practitioners for the evaluation of various neurological disorders. And it is clearly known that language and social skills are the most negatively impacted skills in ASD population. Saccadic eye movement plays a big role in the attention and behavior of an individual which directly affects both language and social skills. Autistic children seem to have different eye movement than non-autistic children. They tend to avoid eye contact and looking at human face while focusing more on geometric shapes. While a typical child doesn’t find any interest in geometric shapes and tend to make more eye contact, and human face perception. So, there is a high possibility that eye movement has some biomarker for ASD which can be spotted early enough to be used as an objective measure for diagnosis. [18]

2 LITERATURE REVIEW

2.1 Using EEG for ASD diagnosis

Ezno, Chiara, Massimo use a complex EEG processing algorithm called MS-ROM/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 I-FAST 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.

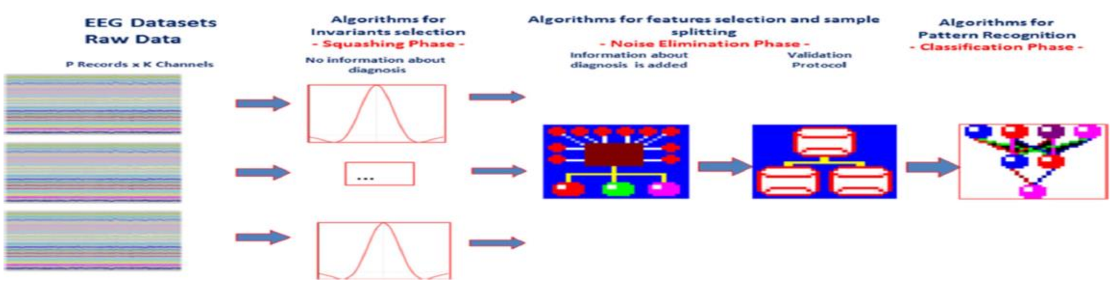

Fig 1. Shows the workflow of the system from raw data to classification

Here, they make comparison between different algorithms like Multi Scale Entropy (MSE) and the Multi Scale Ranked Organizing Maps (MS-ROM). MS-ROM is a new algorithm based on Single Organizing Map Neural Network. This part is completely unsupervised. Secondly they consider the noise elimination stage where unimportant features from the feature vector is removed. Here, 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, 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. Leave one out protocol had the best result with Random Forest with only 1 error. Thy plan to further this study with larger sample to see if their approach is really conclusive. [1]

Bosl, Tierney, Tager-Flusberg, and Nelson 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. [2]

In another study Jamal, Das, Oprescu, Maharatna, Apicella, and Sicca use EEG during facial perception tasks to classify ASD using discriminative analysis and Support Vector Machine. They extracted synchronized patterns of EEG from 128 electrodes from 12 ASD and control individuals. With leave one out cross validation their model's accuracy for differentiating ASD and typical population is 94.7%. Although this is a good result they would like to validate with larger data to confirm the effectiveness of their method. [3]

Abdulhayh, Alafeef, Hadoush, Alomari, and Bashayreh 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. [4]

Dejman, Khadim, and Khorrami use transfer entropy and graph theory with EEG to analyze brain network of Autism. In this study they collected data from 12 high functioning Autistic Youths and 19 healthy control group. EEG signals were collected while the subjects looked at human faces on a screen in a dark room with a sampling frequency of 1KHz. The raw signals are preprocessed to remove noise and transfer entropy is computed. Transfer Entropy quantifies the information transferred between two channels of the EEG. As a first step transfer entropy between all the two pair of channels is calculated. Then they create a graph where channels are the nodes and edges are the connections between channels with transfer entropy above certain threshold. And they use shift test to remove noisy transfer entropy values. Then they perform further computation on the created brain network graph. Mainly they calculate the average degree, total clustering coefficient, average path length, and the longest path length of the brain network. And these estimated values of the graph between ASD and healthy controls are compared using independent sample t-test with permutation to identify the differences. A significant difference was found between the average degrees of ASD and healthy controls. The average degree of healthy controls were much higher than that of ASD. And the authors point to the fact that this aligns with the existing connectivity theory of ASD. That is Autistic brain consists of much less neural connections and that information transfer between nodes in the brain is much difficult. This study shows that lower average degree of the effective connectivity can be a biomarker of Autism. Authors intend to use other graph parameters along with multivariate effective connectivity measures and collect EEG at different types of perception tasks in the future.[5]

Begum, Ravikumar, Vykuntaraju utilize EEG signals to classify Autism, Global developmental delay, and Epilepsy. Unlike most of the other studies that make binary comparisons between Autism and healthy this study compares three different neurological disorder. Here they make both types of multiclass and binary classification. For the data they collect EEG from 10 neurologically impaired and normal subjects of age group between 2 weeks and 15 years using 23 electrodes. The sleep state EEG of children below 5 years was collected whereas awake state EEG was collected for those above 5 years of age. First the raw signals were preprocessed to remove noise like eye movement and other artifacts. Then it was normalized between -1 and 1 which was then filtered using bandpass equiripple filter with sampling frequency of 250 Hz.Then the features were extracted using approximate entropy, sample entropy, and recurrence quantification analysis. Altogether they compute 20 features of approximate entropy, sample entropy, RR, DET, ENTR, energies, and AR coefficients of wavelets. Then they train various machine learning models with 5 fold cross validation. For the multiclass classification between all three they use lda, decision trees, SVM, and KNN. Here there best accuracy was 67.7% for SVM and 59.0% for KNN. For binary classification between Autism and Epilepsy they obtain 82.5% using SVM and 78.1% using KNN. For binary classification between Autism and GDD they obtain 82.2% using SVM and 75.0% using KNN. For classification between GDD and Epilepsy theu obtain 70.6% using SVM and 65% using KNN. The authors mention that the low accuracy for Epilepsy and GDD might be because of the high likelihood of overlap between these two conditions. This proposed method can be done with more subjects to confirm the validity of this method and perhaps increase the accuracy.[6]

Shams, and Rahman classify Autism with EEG using Principal Component Analysis and Neural Network in motor movement and open eyed tasks. In this study they collected data from 6 Autistic and 6 typically developing children between age group 7 and 9 using 8 electrodes. Two different sets of EEG was collected with motor movement and just open eyed tasks. They preprocess the data by normalizing each channel and downsampling the signals from 250 Hz to

83.3 Hz. And it is filtered using Butterworth band pass filter. Then the data is smoothed by averaging. Next, short time fourier transform is performed after which Principal Component Analysis is computed. PCA helps in classification by reducing the dimension and selecting more important features. Then they train a Multilayer Perceptron or a Neural Network to classify the two groups. For testing they use two different protocols. In the first one thy mix all the data together and randomly divide into training and testing. Here the models accuracy of motor tasks for predicting Autism was 100% while the accuracy for typically developing was 85.3%. And for the eye opened tasks the accuracy for Autism was 86% and typically developing was 78.9%. In the second case they train with all the subjects except one test subject which is performed for each subject. For this instance the accuracy with motor tasks for Autism was 100% while for typically developing was 99%. With open eyed tasks the accuracy for Autism is between 40% to 80% and typically developing is around 90%. This study shows that motor tasks are better in detecting Autistic properties than open eyes tasks. Hence the author suggests future works should focus on motor regions for better prediction. [7]

Djemal, AlSharabi, Ibrahim, and Alsuwailem use wavelet, entropy, and Neural Network with EEG to classify Autism. In this study they collected resting state EEG with 16 channels from 10 healthy participants of age group between 9 and 16 and 9 autistic participants of age group 10 to 16 years. Eye artifacts are removed using Independent Component Analysis. And for filtering they make use of elliptic band-pass filter because of its efficiency and efficacy. They produce two set of datasets one of which is comprised of overlapping segment of signal and the other with nonoverlapping segments. These datasets are further processed using wavelet decomposition and entropy function. They use 4-level discrete wavelet transform decomposition to get 4 levels of coefficient D1-D4 along with approximate coefficient A4. After this they compute entropy values of these coefficients. Here they utilize and compare four different entropy functions like log energy entropy , threshold entropy, Renyi entropy, and Shannon entropy. This process from feature extraction to classification is shown below.

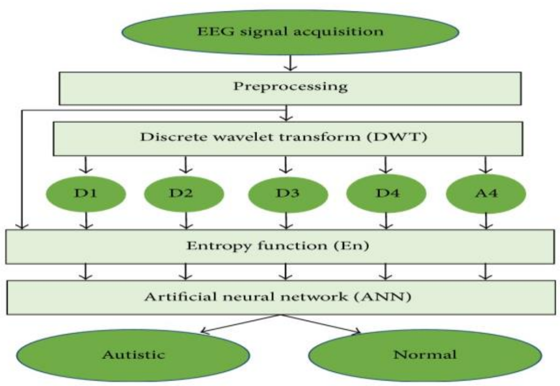

Fig 2. Shows the workflow of this system using DWT with Entropy function.

Then for the classification they train mainly two different models. One using discrete wavelet transform statistical values and the other one using discrete wavelet transform with entropy function. Their neural network consists of 3 layers with just one hidden layer. Hidden layer consists of five nodes with sigmoid function and output layer consist of two nodes with softmax. For evaluation they use 10 fold cross validation. They first experiment with these different cases and select the best one for further optimization. With DWT statistical values the accuracy was the highest of 96% with standard deviation while it was very low for others. DWT with Entropy function had the highest accuracy of 98.4% with Shannon Entropy while the other entropies also had good accuracy above 83%. Then to optimize they use the method with the highest accuracy DWT with Shannon entropy for further analysis. In this step they select the best segment length, best frequency band and best window segmenting technique. For segment length they test from 10 to 180 seconds to find the best segment length to be 50 seconds. For frequency band or wavelet coefficient they test with all the combinations of the dwt coefficients D1,D2, D3, D4, A4. They find the original or the combination of all the 5 coefficients to produce the best result. Comparing non overlapping and overlapping segments they find that overlapping windows produces best result with about 99.7% accuracy. Here, they found the the most optimal method of feature extraction and classification. The authors hope to continue this work with larger datasets and other noise removal algorithms. [8]

Laxmi, and Priya use Cascade forward back propagation neural network and Elman Neural Network with EEG to classify Autism. They collect EEG data using 10 electrodes from 4 normal and 6 autistic children of age group between 6 to 12 years. EEG of various states were collected. Subjects were kept in a relaxed state, made to read and spell flashcards, read and spell from video, and imitate hand movement in a video. Similar to other studies they utilize band pass filter to filter some noise. Then feature extraction was done by using four auto regressive algorithms : AR Burg, AR Modified Covariance, AR Covariance, and AR Yule-Walker Method. After this they train Cascade forward back propogation neural network and Elman Neural Network to classify Autism using all these different feature sets. Cascade forward back propagation neural network is a type of feedforward neural network whereas Elman Neural Network is a type of Recurrent Neural Network. Cascade Neural Network with AR Burg had the highest accuracy of 95.21% for subject 9 and the lowest accuracy of 91.36% for subject 2. Similarly Elman Neural Network with AR Burg had the highest accuracy of 95.63% for subject 9 and the lowest of 92.92% for subject 2. Here the author have presented the analysis of combined EEG from four different tasks with autoregressive feature extraction algorithm with two different Neural Networks.[9]

2.2 Using Eye Tracking for ASD diagnosis

Canavan, Chen, Chen, Valdez, Yager, Lin, and Yin make use of eye movement and demographic feature to classify Autism using various tree based machine learning algorithms. Here, they use existing data from the National Database for Autism Research. This dataset consisted of demographic information, eye gaze coordinates, fixations, diagnosis of Autism or associated risk. There were about 257 subjects out of which 91 were female and 166 were males. For their purpose they create one feature vector for each subject of length 2583 with gaze points, age, gender, and average fixation length. For classification they use Random Regression Forests, C4.5 Decision Trees, and PART algorithm and compare their results to find which one is better for detecting Autism.

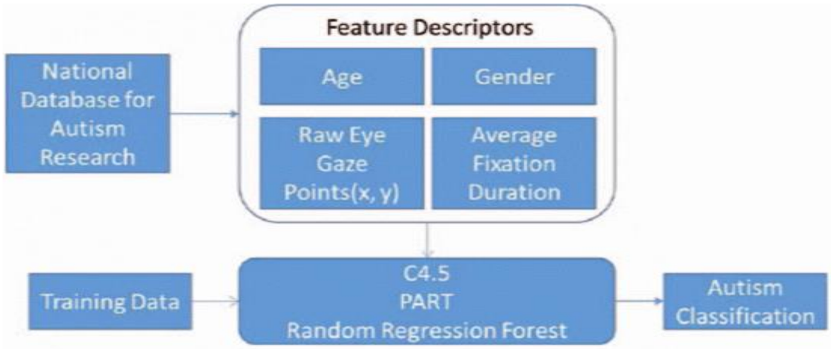

Fig 3. This shows the overview of this system

Random Regression Forests are randomly trained regression trees. C4.5 Decision trees are decision trees based on information entropy. PART known as Projective Adaptive Resonance Theory uses partial decision trees to create rules and filter inputs with matching rules. They make use of 10 fold cross validation to evaluate their models. Also the dataset consisted of subjects above 60 years of age who were diagnosed with Autism which were marked as outliers by the authors. With original data they got 94.16% accuracy from PART, 94.16% accuracy from C4.5, and 91.05% Random Regression Forest. With outliers removed they got 96.2% accuracy from PART, 94.94% accuracy from C4.5, and 93.25 % accuracy from Random Regression Forest. For future work they hope to work with large datasets of children and possibly using deep learning algorithms with combination of eye movement, demographic information and gait. [10]

Torri, Ohtani, and Ishii create an objectivity index to detect Autism using ocular movement. They used eye movement data recorded using camera in front of the PC from 54 Autistic children, and 30 normally developing children from elementary school to high school. They calculated the average of change between frames of the eye movement. Here they use image processing techniques to extract the proper movement features from the camera. First they select the pixel for the subject's pixel, select a reference image called afterimage. Then the pupil is divided to four different sections from the center of the frame image. The number of pixels of the overlapping sections are counted. As images are processed with 30-40 fps the change in pixels of the left and right eye balls are computed. Then they visualize the eye patterns of three of the three groups Autistic, Non Autistic, and not following instruction in separate graph for comparison.

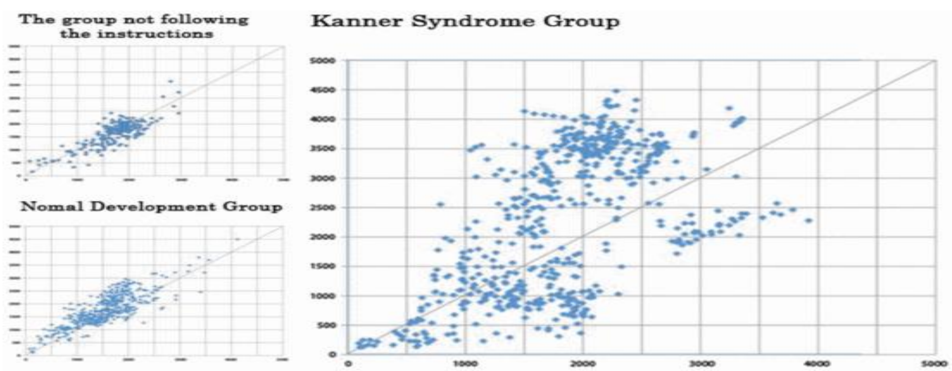

Fig 4. Shows the eye pattern of Autistic on the right. Normal development on the bottom left and other group on the top left.

They found the pattern for Autistic is clearly more scattered than the normal group. By using the normal developing data's density function they calculate the identification border from which they can identify Autism. As this approach works with only PC camera it is very cost effective and easy to use for medical practitioners. This even simplifies testing as data can be collected by global participants using a web application. [11]

Alie, Mahoor, Mattson, Anderson, and Messinger 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. [12]

In another study Vu, tran and colleagues use Eye tracking with varying visual stimulus and K Nearest Neighbor algorithm to differentiate Autism from non Autistic. They make use of Tobii EyeX tracker with three different type of images for 5 seconds each. There were total 32 subjects between ages of 2 and 10 out of which 16 were diagnosed with Autism while the other half were normally developing. The groups of stimulus or images were divided as social scene, human face, and object. Social scene comprised of 4 images of social events. Human face group comprised of 4 images of sad,happy,blurred, normal type faces. Object group had 4 images of different food, toys, animals, and a drawing.

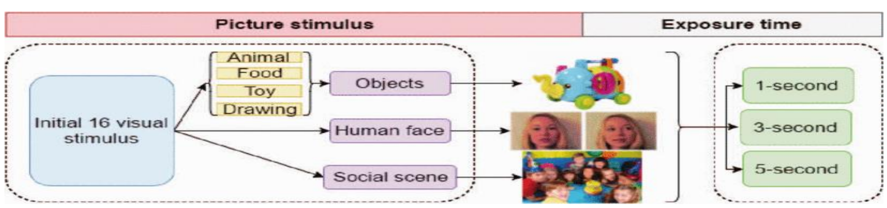

Fig 5. Shows how the visual stimulus and time are divided for the experiment

The total exposure time of 5 seconds was divided into 3 interval times of 1 second, 3 seconds and full 5 seconds. This way there were total six instances of collections to train and test for. They compute the Wassertein Distance between gaze distributions which was then analyzed by a kNN. Using this approach they had the highest accuracy of 98.24% with Social Scene at 5 seconds. And their lowest accuracy was 78.58% with Object scene at 3 seconds. The accuracy was consistent for both ASD and typically developing. Here, they found the most effective stimulus and exposure time for identifying Autism with visual cues. This really gives a way for future research to focus on Social stimulus with eye tracking to come up with better diagnostic means for ASD. [13]

Liu, Yu, Raj, Yi, Zou, and Li 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. [14]

In another study at University of Minnesota Jiang, and Zhao 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.

$$D = \frac{1}{1 + e^{-I/\sigma_I}}$$

where *I* = Autism Fixation Map - Control Fixation Map

The above formula was used to compute the difference of fixation maps. 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 downsampled 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. [15]

3 RESEARCH GOAL

Identifying ASD is a very complex yet important task. Current techniques in practice 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 however researchers have long found that it can be seen as early as 1 year. 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. An objective measure like MRI, CT, MEG, blood testing, EEG or Eye movement is needed to make this come to life. Through these measures we can get the diagnosis right away. Medical Imaging and blood testing 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 by many people like in remote areas. So, there is a need for more investigation in the diagnosis of ASD using EEG and Eye movement. There has been various work in this area but there is still a lot of room for improvement. Specially all the work so far has used isolated EEG and Eye movement for identifying ASD. There has not been any work yet that has combined and compared both EEG and Eye Movement for classifying ASD.

3.1 Goals

The aim of this research is to shed light on the identification of Autism Spectrum Disorder using both EEG and Eye Tracking. The goals are as follows:

A. Primary goal is to analyse the classification of 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 Movement 3) Combination of EEG and Eye Movement. Comparison of the classification performance between EEG, Eye Movement and combination of both EEG and Eye Movement can potentially result in finding the superior features. Classification using these different features will tell us which one of these feature set is a better indicator of ASD. This can in turn show which signal source have more details about ASD as the top performing signal most likely has more of the unique data points and pattern of ASD. Similarly, the least performing signals have less of the data points and patterns relating to ASD. In other words this research can be looked as a feature selection process between EEG, Eye Movement and combination for the diagnosis of ASD. So, in the future work more investigation on the better working singal sources can be done to develop clinical systems.

B. The secondary goal is to compare various machine learning algorithms like SVM, Deep Neural Network, Logistic, and Naive Bayes for the classification purpose. Comparing these different models may also reveal some information about the features themselves. Like some models are better for linearly separable data points whereas others like Multilayer Perceptron is able to work with non linear data points. Or at least we will know which models are better suited for EEG, Eye movement and combined features.

C. In addition the finding of this work may not just be limited to ASD as we are working with biophysical features. Conditions like ADHD, and other learning disabilities can also share similar comparative patterns for different features.

4 DATA

Data from a previous study “Reduced Temporal-Central EEG Alpha Coherence during Joint Attention Perception in Adolescents with Autism Spectrum Disorder” was used for this thesis. As this was the only study so far that collected both eye movement and EEG of both ASD and control group. For ASD they collected high-functioning adolescents from Nova Southeastern University Center for Autism. For the control group they collected data from typically developing adolescents of the same age group from Miami-Dade public schools. Initially there were 52 participants consisting of 24 ASD and 28 control. Some of the participant’s data was not used in the study because some were misdiagnosed with ASD, or didn’t meet the some of the required scores and missing EEG data. In the end there were total 33 participants. They retrieved two types of Eye Movement data whereas just one for EEG.

4.1 EEG Data

EEG was collected using a 128-channel Electrical Geodesics Inc. The studies aim was to analyse the cortical connectivity during joint attention by examining ASD and typically developing groups. Joint attention is the ability to maintain proper eye coordination in social interaction like making eye contact during speaking etc.

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 thesis. Note that Dr. Jaimes study only used the final 33 participants data for analysis.

4.2 Eye Test Data

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.

4.3 Eye Movement Data

A Tobi X50 eye tracker was used to collect the eye movement data. Visual fixation times were calculated using the eye trackers proprietary software.

For this thesis the eye movement data consists of the computed fixation time.

And the final eye data comprises of eye test data concatenated with the eye movement data.

There were a total of 36 participants Eye data.

All the implementation

4.4 Training and Testing

For classification I separated the data into about 80% training and 20% testing. Testing data was strictly used only for testing while training for training only. 27 out of 33 were chosen for training while the rest 7 were used for testing. This was done manually by selecting the right data to separate for testing. Amongst the 7 test data 4 were diagnosed with ASD while the other 3 were not.

5 EEG PREPROCESSING

EEG raw data was preprocessed using Makoto Preprocessing pipeline.

5.1 Makoto Preprocessing Pipeline

This data was preprocessed using EEGLab. Makoto Pipeline is the most recommended EEG preprocessing pipeline using EEGLab. It was the pipeline firstly recommended by Makoto Miyakoshi hence its named after. This pipeline usually consists of the following steps

1. Converting data points to double precision
2. Downsampling only if necessary
3. High-pass filter at 1-Hz (for ICA, ASR, and CleanLine)
4. Import channel info
5. Remove line noise using CleanLine
6. Apply clean_rawdata() to reject bad channels and correct continuous data using Artifact Subspace Reconstruction (ASR).
7. Interpolate all the removed channels
8. Re-reference the data to average
9. Run AMICA using calculated data rank with 'pcakeep' option (or runica() using 'pca' option)
10. Estimate single equivalent current dipoles
11. Search for and estimate symmetrically constrained bilateral dipoles
12. Create STUDY with no STUDY.design just for IC rejection (this greatly saves your time--see this page for how to do it)
13. Epoch IC-rejected data to -1 to 2 sec to event onset
14. Create final STUDY specifying full STUDY.design

[21]

The main purpose of this pipeline is to remove noise or artifacts -- mainly eye blink artifacts -- from the EEG data.

6 FEATURE EXTRACTION

6.1 EEG

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 I 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 shannon entropy function over windows of signals.

1. Statistical Features

    a. Mean

    Here the mean of 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 figure below.

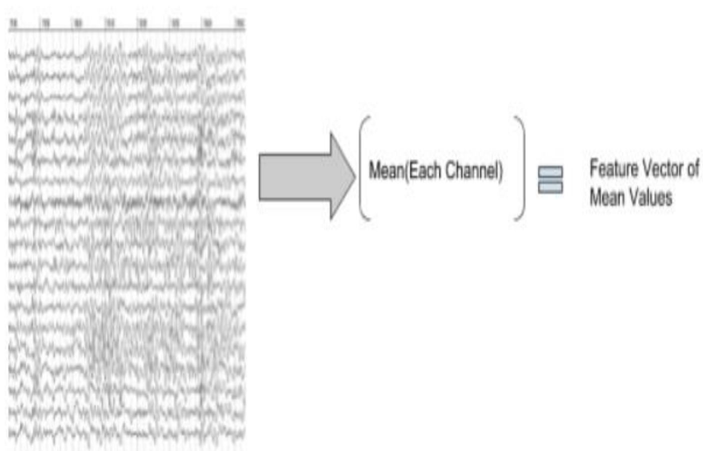

Fig 6. Raw EEG signal with 128 channels (2D matrix)-> Compute Mean of each of the channel and combine mean of all the channels -> Vector with mean values of each channel

b. FFT

Discrete Fast Fourier Transform is computed using the formula below.

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

I use an open source python library called ***SciPy*** to calculate the FFT.

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.

Initial Feature Matrix of a subject

0.1783269101	5.3477088859	0.178343793	5.3475004223	0.1783604307	
0.1655341136	5.3406293864	0.1655499051	5.3402707063	0.1655649181	
0.347215635	5.3149732262	0.347215635	5.3149732262	0.3471614871	...
0.0264205348	5.3570682653	0.0265605849	5.3569125702	0.0264638036	
0.183816315	5.3406464867	0.1838385347	5.3400792717	0.183816315	...
0.1614730338	5.3414577663	0.1614730338	5.3414577663	0.1614730338	
0.0743148651	5.3534797759	0.0743148651	5.3534797759	0.0743741934	...
0.2152908486	5.3158405987	0.2152908486	5.3158405987	0.2150407211	...
0.9979478592	4.888887039	0.9979478592	4.888887039	0.9979478592	
0.5263203611	5.1657623313	0.5236650479	5.15763225	0.5250101208	...
0.1091793638	5.3455155123	0.1091793638	5.3455155123	0.1092399887	...
0.2094936626	5.3372109621	0.209514211	5.3359521108	0.2094936626	
...					

Fig. 7 Each coloum represents a channel and each row represents a time.

Subject DFFT Matrix

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

Fig 8. Represents the FFT matrix of one subject. Each coloum represents a channel and rows represent a window.

Channel1 = $\begin{bmatrix} FFTWindow1 \\ FFTWindow2 \\ FFTWindow3 \\ FFTWindow4 \\ \dots \end{bmatrix}$

$$\begin{bmatrix} SDFFT \\ MeanFFT \end{bmatrix}$$

Fig 9. Standard deviation and mean is computed for each channel separately

Final Feature Vector

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

Fig 10. Each feature vector consists of the mean and standard deviation of the windows of each channel.

c. Standard Deviation

Here the standard deviation of 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.

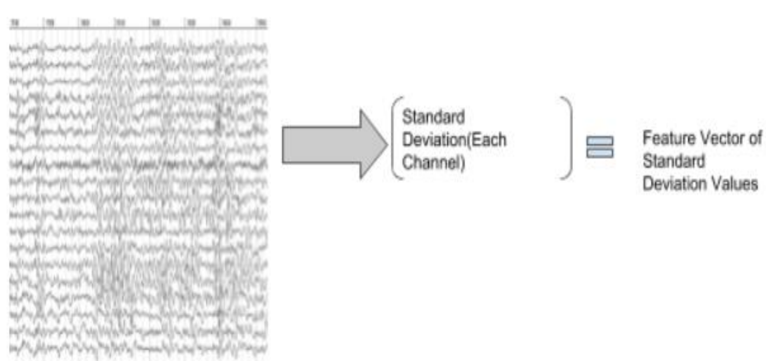

Fig 11. Raw EEG signal with 128 channels (2D matrix)-> Compute Standard Deviation of each of the channel and combine standard deviation of all the channels -> Vector with standard deviation values of each channel

2. Entropy

    a. Shannon Entropy

    Shannon Entropy “is the average rate at which information is produced by a stochastic source of data” [22]

    It is computed using the formula below.

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

I use an open source python library called ***pyentrp*** to get the entropy values.

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. After this standard deviation and mean of the entropy 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 entropy values of each channel.

Below are the matrices and vectors of features.

Initial Feature Matrix of a subject

0.1783269101	5.3477088859	0.178343793	5.3475004223	0.1783604307	
0.1655341136	5.3406293864	0.1655499051	5.3402707063	0.1655649181	
0.347215635	5.3149732262	0.347215635	5.3149732262	0.3471614871	…
0.0264205348	5.3570682653	0.0265605849	5.3569125702	0.0264638036	…
0.183816315	5.3406464867	0.1838385347	5.3400792717	0.183816315	…
0.1614730338	5.3414577663	0.1614730338	5.3414577663	0.1614730338	…
0.0743148651	5.3534797759	0.0743148651	5.3534797759	0.0743741934	…
0.2152908486	5.3158405987	0.2152908486	5.3158405987	0.2150407211	…
0.9979478592	4.888887039	0.9979478592	4.888887039	0.9979478592	
0.5263203611	5.1657623313	0.5236650479	5.15763225	0.5250101208	
0.1091793638	5.3455155123	0.1091793638	5.3455155123	0.1092399887	
0.2094936626	5.3372109621	0.209514211	5.3359521108	0.2094936626	

…

Fig. 12 Each coloum represents a channel and each row represents a time.

Subject Entropy Matrix

$$\begin{bmatrix} EntropyWindow1Channel1 & EntropyWindow1Channel2 & \dots \\ EntropyWindow2Channel1 & EntropyWindow2Channel2 & \dots \\ EntropyWindow3Channel1 & EntropyWindow3Channel2 & \dots \\ EntropyWindow4Channel1 & EntropyWindow4Channel2 & \dots \\ \dots & & \end{bmatrix}$$

Fig 13. Represents the entropy matrix of one subject. Each coloum represents a channel and rows represent a window.

$$\text{Channel1} = \begin{bmatrix} EntropyWindow1 \\ EntropyWindow2 \\ EntropyWindow3 \\ EntropyWindow4 \\ \dots \end{bmatrix} \Rightarrow \begin{bmatrix} SDEntropy \\ MeanEntropy \end{bmatrix}$$

Fig 14. Standard deviation and mean is computed for each channel separately

Final Feature Vector

$$\begin{bmatrix} SDEntropyChannel1 \\ MeanEntropyChannel1 \\ SDEntropyChannel2 \\ MeanEntropyChannel2 \\ \dots \end{bmatrix}$$

Fig 15. Each feature vector consists of the mean and standard deviation of the windows of each channel.

6.2 Eye Data

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

7 METHODOLOGY

All the implementation has been done in python using keras with Tensorflow. Classification has been done in many different ways. As a first step we are simply classifying if a participant has ASD or not. As this is a binary classification problem I expect to be simpler than a multiclass where we would be predicting multiple conditions. Multiclass classification is one of the future works. 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 to predict from. Hence this became a quiet complex analysis with many different models.

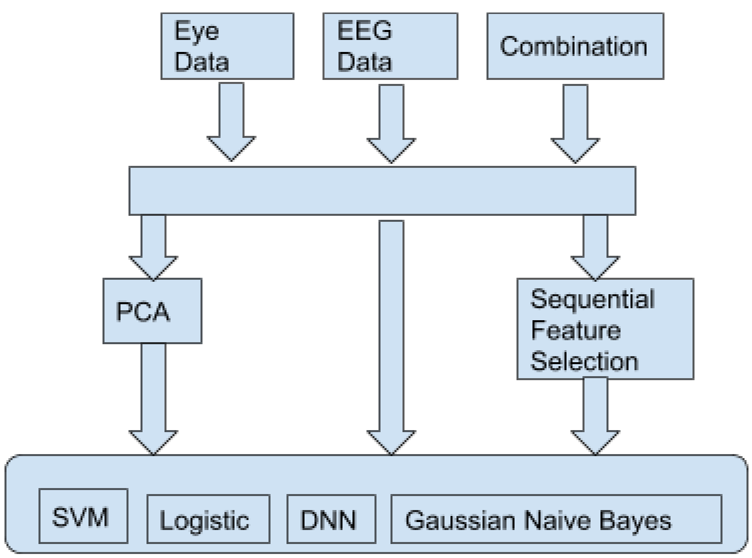

Figure 16. Methodology of classification

Mainly there are three types of classification :

A. Classification using EEG

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

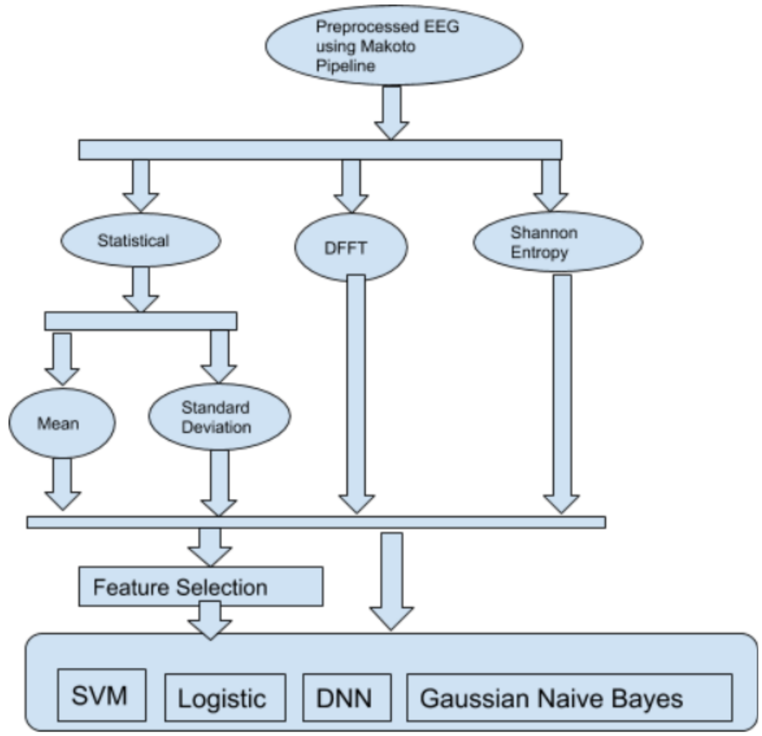

Fig 17. EEG classification pipeline

As shown in the image there are mainly three different feature set -- Entropy features, FFT and statistical features -- to predict from. And there are two different types of statistical features which are simply mean, and standard deviation. 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. So, overall there are 16 different models. Four Logistic regression models for mean, dfft, standard deviation, and shannon entropy. Four SVM models for mean, dfft, standard deviation and shannon entropy. Four Deep Neural Network models for mean, dfft, standard deviation, and shannon entropy. Four

Gaussian Naive Bayes models for mean, dfft, standard deviation and shannon entropy. 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.

A.1 Deep Neural Network

Deep neural network with five hidden layers with sigmoid activation function was used. First Input layer had numberFeatures + 1 neurons while the rest had numberFeatures neurons. For optimization categorical cross entropy for loss and Adamax optimizer was used. Below is the architecture of the network.

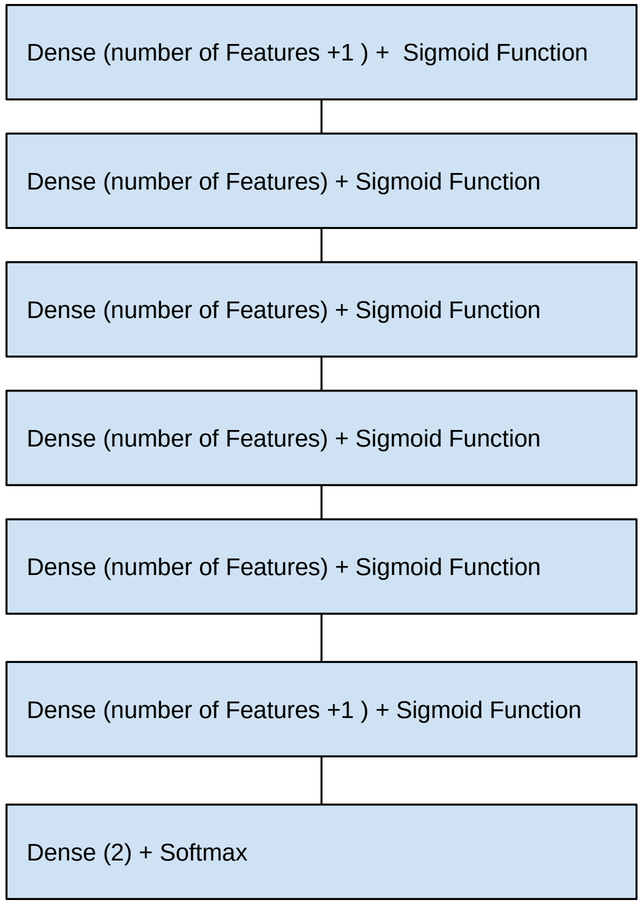

Fig 18. Architecture of DNN using EEG data

B. Classification using Eye Movement

Below is the high level overview of the classification pipeline.

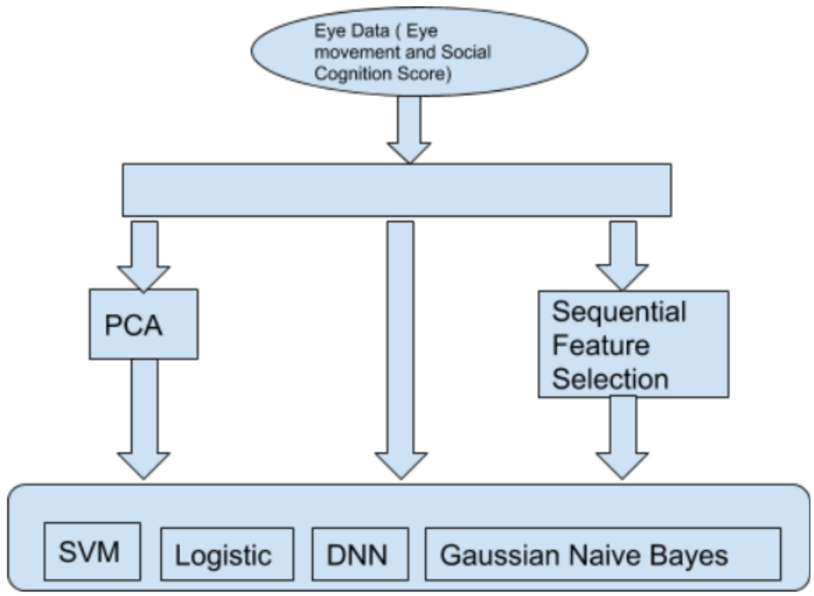

Fig 19. Classification Pipeline using Eye data

As shown in the image there are 4 classifiers working on the Eye data. Eye data comprises of the Eye movement fixation times and the Eye test data.

B.1 Deep Neural Network

Deep neural network with four hidden layers with sigmoid activation function was used. First Input layer had numberFeatures + 1 neurons while the rest had numberFeatures neurons. For optimization categorical cross entropy for loss and Adamax optimizer was used. Below is the architecture of the network.

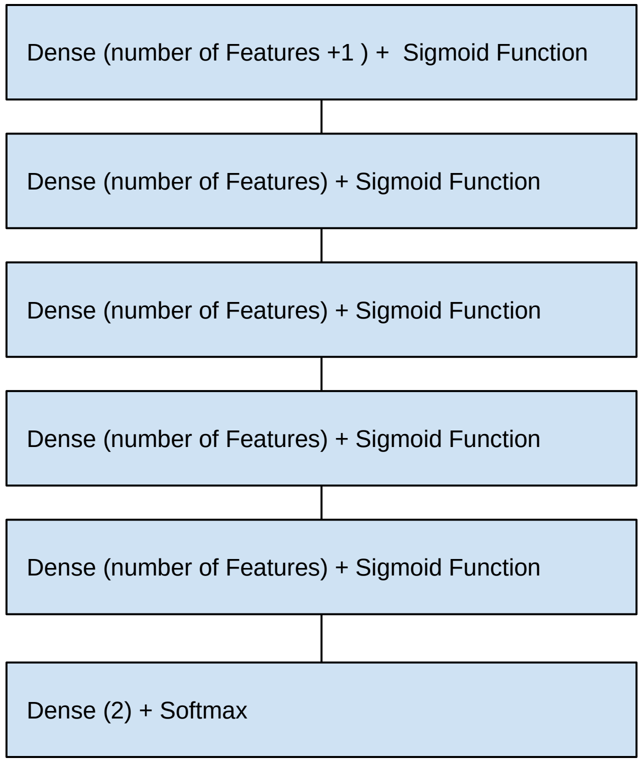

Fig 20. Architecture of DNN using Eye data

C. Classification using Combination of EEG and Eye Data

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

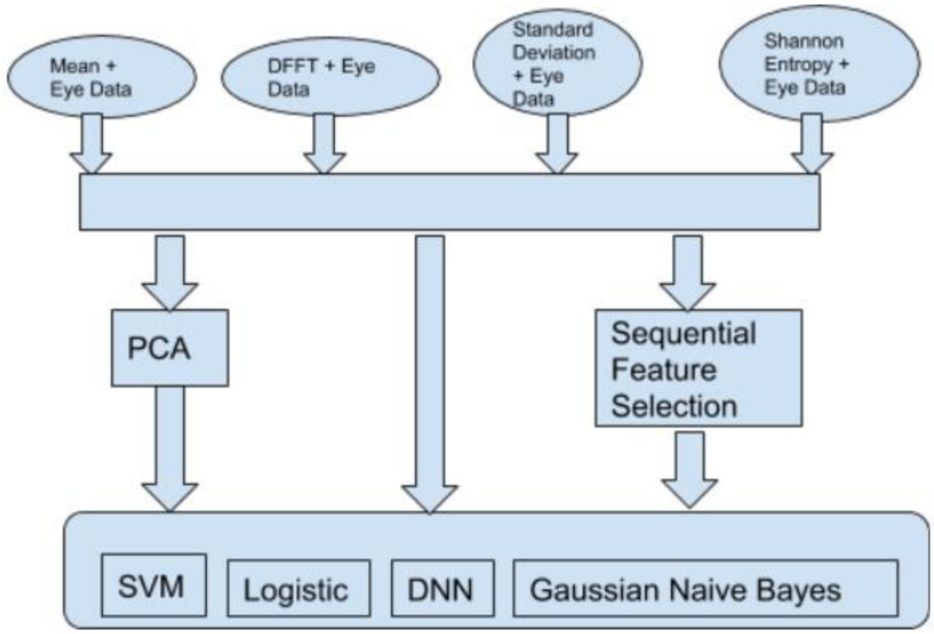

Fig 21. Classification Pipeline using combination of EEG and Eye Data

As shown in the figure there are a total of 16 different models. Four SVM models for mean with Eye data, DFFT with Eye data, Standard Deviation with Eye data, and Shannon Entropy with Eye data. Four Logistic regression models for mean with Eye data, DFFT with Eye data, Standard deviation with Eye data, and Shannon Entropy with Eye data. Four Deep Neural Network models for mean with Eye data, DFFT with Eye data, Standard deviation with Eye data, and Shannon Entropy with Eye data. Four Gaussian Naive Bayes models for mean with Eye data, DFFT with Eye data, Standard deviation with Eye data, and Shannon Entropy with Eye data.

C.1 Deep Neural Network

Deep neural network with no hidden layer was used. First Input layer had numberFeatures + 1 neurons with sigmoid function. For final layer softmax with 2 neurons was used. For optimization categorical cross entropy for loss and Adamax optimizer was used. Below is the architecture of the network.

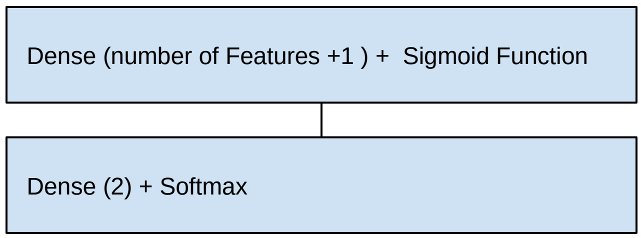

Fig 22. Architecture of DNN using combination of EEG and Eye features

8 RESULTS

The three datasets has been analyzed in different ways. The results are as follows.

8.1 With Statistical Measure

A. Using Mean

- SVM

With and without PCA

Loss: 19.7
Accuracy: 0.428

Confusion matrix

Predicted	ASD	Typical	All
Actual			
ASD	0	4	4
Typical	0	3	3
All	0	7	7

Using SVM with mean we can see that all the Typical Development are classified correctly whereas all the ASD are misclassified.

- Logistic Regression

With and without PCA

ACC: 0.571428571429

Confusion matrix

Predicted	ASD	Typical	All
Actual			
ASD	2	2	4
Typical	1	2	3
All	3	4	7

Using Logistic with mean values the two out of two ASD were classified correctly. While two out of three non ASD were classified correctly.

- <u>Deep Neural Network</u>

  With and without PCA

  Loss: .90
  Accuracy: .57
  Confusion Matrix

Predicted	ASD	Typical	All
Actual			
ASD	2	2	4
Typical	1	2	3
All	3	4	7

Using DNN with Mean values has the same results as Logistic regression.

- <u>Gaussian Naive Bayes</u>

With and without PCA

Acc: 0.57

Confusion Matrix

Predicted	ASD	Typical	All
Actual			
ASD	4	0	4
Typical	3	0	3

Using Mean values with Gaussian Naive Bayes predicted four out of four correctly for ASD. While for non ASD none were predicted correctly.

B. Using Standard Deviation

- <u>SVM</u>

Without PCA
Loss: 19.736
Accuracy: 0.428

Confusion Matrix

Predicted	ASD	Typical	All
Actual			
ASD	0	4	4
Typical	0	3	3
All	0	7	7

Using SVM with standard deviation features predicted all the ASD subjects correctly. While all the non ASD subjects were predicted incorrectly.

With PCA

Loss: 14.8
Accuracy: 0.57

Predicted	ASD	Typical	All
Actual			
ASD	1	3	4
Typical	0	3	3
All	1	6	7

Using PCA the overall accuracy got better from 42% to 57%. But just one out of four ASD subjects were classified correctly.

- <u>Logistic Regression</u>

With PCA
ACC: 0.57

Predicted	ASD	Typical	All
Actual			
ASD	2	2	4
Typical	1	2	3
All	3	4	7

Using Logistic with Standard deviation has an overall 57% accuracy. Two out of four ASD were classified correctly. And two out of three typically developing were classified correctly.

Without PCA
Accuracy: 0.71

Predicted	ASD	Typical	All
Actual			
ASD	3	1	4
Typical	1	2	3
All	3	4	7

Using PCA we can see that the results have shifted from 57% to 71%. Only one ASD subject out of four was misclassified. For typically developing it is the same as without using PCA.

- <u>Deep Neural Network</u>

With using PCA

Loss: 0.48
Accuracy: 0.85

Confusion Matrix

Predicted	ASD	Typical	All
Actual			
ASD	4	0	4
Typical	1	2	3
All	5	2	7

Using DNN with standard deviation features resulted in overall 85% accuracy. Four out of four ASD participants were classified correctly. And two out of three typically developing were classified incorrectly. This model outperformed all the other models with Standard deviation features. This might be because the data was not very linear. And DNN was able to learn those hidden non linear patterns.

Without using PCA

Loss: 0.718
Accuracy: 0.428

Predicted	ASD	Typical	All
Actual			
ASD	1	3	4
Typical	1	2	3
All	2	5	7

Without using PCA DNNs accuracy dropped from 85% to 42%.

- <u>Gaussian Naive Bayes</u>

  Both with and without PCA
  Num Components: 27

  Accuracy: 0.714

Predicted	ASD	Typical	All
Actual			
ASD	4	0	4
Typical	2	1	3
All	6	1	7

Using Gaussian Naive Bayes with Standard deviation features resulted in overall 71% accuracy. Here all the ASD participants were predicted correctly. Whereas only one out of three typically developing were predicted correctly.

8.2 With Shannon Entropy

Using shannon Entropy I used two set of features. One set of features were computed for about half of the original length of the EEG whereas the other set used the entire length of the EEG. This happened by accident but the results came out to be comparable. And it brings about an interesting analysis between the time length of the EEG data.

A. Using half of the EEG entropy features

- SVM

With and without PCA
Loss: 19.73
Accuracy: 0.428

Predicted	ASD	Typical	All
Actual			
ASD	0	4	4
Typical	0	3	3
All	0	7	7

The overall accuracy using SVM is around 42%. All the ASD participants were classified correctly whereas all the typically developing were classified incorrectly

- Logistic

  With PCA
  Num Components = 27

  Accuracy: 0.57

Predicted	ASD	Typical	All
Actual			
ASD	3	1	4
Typical	2	1	3
All	5	2	7

Using logistic regression with PCA produces an overall accuracy of 57%. Three out of the four ASD participants were classified correctly whereas one out of the three typically developing were classified correctly.

Without PCA

Accuracy: 0.71

Predicted	ASD	Typical	All
Actual			
ASD	2	2	4
Typical	0	3	3
All	2	5	7

Without using PCA the overall accuracy shifted from 57% to 71%. Two out of the four ASD participants were classified correctly. While all the typically developing participants were classified correctly.

- Deep Neural Network

  With and without PCA

  Loss: 0.71
  Accuracy: 0.42

Predicted	ASD	Typical	All
Actual			
ASD	0	4	4
Typical	0	3	3
All	0	7	7

Using DNN with half of the entropy features produced an accuracy of 42%. All the ASD participants were misclassified. Whereas all the typically developing were classified correctly.

- Gaussian Naive Bayes

With PCA

Accuracy: 0.428

Predicted	ASD	Typical	All
Actual			
ASD	0	4	4
Typical	0	3	3
All	0	7	7

Using Naive Bayes with PCA performed the same as DNN with an overall accuracy of 42%.

Without PCA

Accuracy: 0.57

Predicted	ASD	Typical	All
Actual			
ASD	4	0	4
Typical	3	0	3
All	7	0	7

Using Naive Bayes without PCA shifted the accuracy from 42% to 57%. All the ASD participants were classified correctly. While all the typically developing participants classified incorrectly.

B. Using all the features

SVM

With PCA

Loss: 19.73
Accuracy: 0.428

Predicted	ASD	Typical	All
Actual			
ASD	0	4	4
Typical	0	3	3
All	0	7	7

Using all the entropy features with SVM has the same performance as with half the features.

Without PCA
Num Components = 258

Loss: 9.86
Accuracy: 0.71

Predicted	ASD	Typical	All
Actual			
ASD	2	2	4
Typical	0	3	3
All	0	5	7

Without using PCA the overall accuracy shifted from 42% to 71%. This result is better than that with half the features.

- <u>Logistic Regression</u>

With PCA

Num Components = 27
Accuracy: 0.28

Predicted	ASD	Typical	All
Actual			
ASD	1	3	4
Typical	2	1	3
All	3	4	7

Using logistic regression with PCA for all the entropy features produced overall accuracy of 28%. This model performed worse than with half the features and all other models.

Without PCA

Num Components = 258
ACC: 0.57

Predicted	ASD	Typical	All
Actual			
ASD	3	1	4
Typical	2	1	3
All	5	2	7

Without PCA the accuracy shifted from 28% to 57%. This result is also worse than that with half the features. With half the features without PCA the accuracy for logistic regression was about 71%.

- Deep Neural Network

With PCA
Num Components = 27

Loss: 0.68
Accuracy: 0.57

Predicted	ASD	Typical	All
Actual			
ASD	4	0	4
Typical	3	0	3
All	0	7	7

Using DNN with PCA produced an overall accuracy of 57%. With half the features the accuracy was only 42%. Here, all the ASD participants were classified correctly. While all the typical developing were misclassified.

Without PCA
Num Components = 258
Loss: 0.93

Accuracy: 0.42

Predicted	ASD	Typical	All
Actual			
ASD	3	1	4
Typical	3	0	3
All	6	1	7

Without PCA the performance got worse and is the same as with half the features.

- <u>Gaussian Naive Bayes</u>

With PCA
Num Components = 27

Acc: 0.57

Predicted	ASD	Typical	All
Actual			
ASD	4	0	4
Typical	3	0	3
All	7	0	7

Using Gaussian Naive Bayes with PCA the overall accuracy is 57%. With half the features the accuracy was only 42%. Here, all the ASD participants were classified correctly whereas all the typically developing were misclassified.

Without PCA

Accuracy: 0.57

Predicted	ASD	Typical	All
Actual			
ASD	1	3	4
Typical	0	3	3
All	1	6	7

Without using PCA the overall accuracy is the same as with PCA. However, the actual results classification varies. Here only one out of the four ASD participants were classified correctly whereas all the typically developing were correctly classified. Also the overall accuracy with only half the entropy features is the same with varying results. However the actual classification with half the features is much better with all correct predictions.

8.3 With DFFT

- <u>SVM</u>

With and without PCA

Loss: 19.73
Accuracy: 0.42

Predicted	ASD	Typical	All
Actual			
ASD	0	4	4
Typical	0	3	3
All	0	7	7

Using SVM with DFFT EEG the overall accuracy is 42%. All the ASD participants were misclassified whereas all the typically developing participants were correctly classified.

- DNN

With and Without PCA
Accuracy: 0.57
Loss: 0.428

Predicted	ASD	Typical	All
Actual			
ASD	0	4	4
Typical	0	3	3
All	0	7	7

In this case the accuracy is low and the classification is poor. All the ASD participants were misclassified whereas all the controls were classified correctly.

- Logistic Regression

Without PCA
Accuracy: 0.42

Predicted	ASD	Typical	All
Actual			
ASD	2	2	4
Typical	2	1	3
All	4	3	7

Using Logistic Regression without PCA the overall accuracy is 42%. Two out of the four ASD participants were correctly classified. And one out of three typically developing were classified correctly.

With PCA
Accuracy: 0.71

Predicted	ASD	Typical	All
Actual			
ASD	3	1	4
Typical	1	2	3
All	4	3	7

With PCA the accuracy shifts to 71%.

- <u>Gaussian Naive Bayes</u>

Without PCA
Accuracy: 0.57

Predicted	ASD	Typical	All
Actual			
ASD	3	1	4
Typical	2	1	3
All	5	2	7

With PCA
Accuracy: 0.428

Predicted	ASD	Typical	All
Actual			
ASD	2	2	4
Typical	2	1	3
All	4	3	7

8.4 With Eye Data

The following models are using the entire eye data that includes the eye fixation times, social cognition scores from eye test, age, gender, and IQ scores.

- SVM

  With and without PCA

  Loss: 14.8

  Accuracy: 0.57

Predicted	ASD	Typical	All
Actual			
ASD	1	3	4
Typical	0	3	3
All	1	6	7

Using SVM with eye data the overall accuracy was around 57%. One out of the four ASD subjects were classified correctly whereas all the typically developing participants were classified correctly.

- <u>Logistic Regression</u>

  With and without PCA

  Accuracy: 0.85

Predicted	ASD	Typical	All
Actual			
ASD	4	0	4
Typical	1	2	3
All	5	2	7

Using logistic regression with eye data had about 85% accuracy. Here all the ASD participants were classified correctly whereas only two out of three typically developing participants were classified correctly.

- <u>Deep Neural Network</u>

  With PCA

  Loss: 0.569
  Accuracy: 0.71

Predicted	ASD	Typical	All
Actual			
ASD	4	0	4
Typical	2	1	3
All	6	1	7

Using DNN with Eye data along with PCA produced an accuracy of about 71%. Here all the ASD participants were classified correctly. And only one out of three typically developing participants were classified correctly.

Without PCA

Loss: 0.86
Accuracy: 0.57

Predicted	ASD	Typical	All
Actual			
ASD	3	1	4
Typical	2	1	3
All	5	2	7

Using DNN with Eye data along without PCA produced an accuracy of about 57%. Here three out of four ASD participants were classified correctly. And only one out of three typically developing participants were classified correctly.

- <u>Gaussian Naive Bayes</u>

Without PCA
Accuracy: 1.0

Predicted	ASD	Typical	All
Actual			
ASD	4	0	4
Typical	0	3	3
All	4	3	7

Using Gaussian Naive Bayes without PCA gave a perfect result. All the participants in the test set were classified correctly.

With PCA
Accuracy: 0.857

Predicted	ASD	Typical	All
Actual			
ASD	4	0	4
Typical	1	2	3
All	5	2	7

Using Gaussian Naive Bayes without PCA gave an accuracy of about 85%. All the ASD participants were classified correctly. And two out of three typically developing participants were classified correctly. It is interesting that the model performed better without PCA than with PCA. Intuitively, models should generally do better with PCA as PCA removes the redundant data which don't help in classification. This means that there are data points that PCA removes that are meaningful to some level. None of the other algorithms except Naive Bayes does better without PCA. This also hints that PCA is in fact working well for other machine learning algorithms. But, Naive Bayes is the exception that is able to learn from non discriminant data points.

8.5 With Combination of EEG and Eye Data

These are the results of using the combination of both EEG and Eye data.

A. Mean with Eye Movement

- SVM

Both with and without PCA

Loss: 19.73
Accuracy: 0.428

Predicted	ASD	Typical	All
Actual			
ASD	0	4	4
Typical	0	3	3
All	0	7	7

Using SVM with Mean EEG and Eye data produces an accuracy of about 42%. All the ASD participants were classified incorrectly. Whereas all the typically developing participants were classified incorrectly. This combination model performed worse than the one with only Eye data but the same as the one with only Mean EEG data. Although the combined data had all the eye data that it could not still use those data to differentiate classes. This might be because that the combined data is more non-linear.

- <u>Logistic Regression</u>

Both With and without PCA
Num Components = 12, 159

Accuracy: 0.85

Predicted	ASD	Typical	All
Actual			
ASD	4	0	4
Typical	1	2	3
All	5	2	7

Using Logistic regression with Mean EEG and Eye data produces and accuracy of about 85%. All the ASD participants were classified correctly. Whereas two out of the three typically developing were classified correctly. This combined model performed better than the one with only Mean EEG data but the same as that with only Eye data.

- <u>Deep Neural Network</u>

With PCA
Num Components = 12

Loss: 0.66
Accuracy: 0.71

Predicted	ASD	Typical	All
Actual			
ASD	2	2	4
Typical	0	3	3
All	2	5	7

Using DNN with mean eeg data and eye data produces an accuracy of about 71%. Two out of four ASD participants were classified correctly. Whereas all the typically developing participants were classified correctly.

Without PCA
Num Components = 159

Loss: 0.66
Accuracy: 0.42

Predicted	ASD	Typical	All
Actual			
ASD	2	2	4
Typical	2	1	3
All	4	3	7

Without PCA the combined mean and eye data produces an accuracy of about 442%. Two out of four ASD participants were classified correctly. And only one out of the three typically developing participants were classified correctly. This combined model performed worse than both with only Eye data and only Mean EEG data. It is an indicator of the curse of dimensionality. As this combined dataset has more features than both other feature sets.

- <u>Gaussian Naive Bayes</u>

With PCA
Number of component = 24
Accuracy: 1.0

Predicted	ASD	Typical	All
Actual			
ASD	4	0	4
Typical	0	3	3
All	4	3	7

Using Gaussian Naive Bayes with mean eeg data and eye data produces a 100% accuracy with the test data. All the participants are classified correctly. The performance of this model is better than both with only eye data and only mean eeg data.

Without PCA

Number of component = 159
Accuracy: 0.85

Predicted	ASD	Typical	All
Actual			
ASD	3	1	4
Typical	0	3	3
All	3	4	7

Without PCA the overall accuracy is about 85%. Three out the four ASD participants are classified correctly. Whereas all the typically developing participants are classified correctly. The performance of this model is worse than only eye data but better than only mean EEG data.

B. Standard Deviation with Eye Movement

- SVM

Both with and without PCA

Loss: 19.73
Accuracy: 0.428

Predicted	ASD	Typical	All
Actual			
ASD	0	4	4
Typical	0	3	3
All	0	7	7

Using SVM with standard deviation EEG data and eye data produces an overall accuracy of 42%. All the ASD participants were misclassified. Whereas all the typically developing participants were correctly classified. The performance of this combined model is the same as that of the SD EEG data but worse than that of only eye data.

- Logistic Regression

With PCA
Num Components = 15

Accuracy: 1.0

Predicted	ASD	Typical	All
Actual			
ASD	4	0	4
Typical	0	3	3
All	4	3	7

Logistic regression with standard deviation eeg data and eye data produces an accuracy of 100%. All the participants are classified correctly. This model outperforms both the models with only eye data and only standard deviation eye data.

Without PCA
Number of component = 159

Accuracy: 0.85

Predicted	ASD	Typical	All
Actual			
ASD	4	0	4
Typical	1	2	3
All	4	3	7

Without using PCA Logistic regression with standard deviation eeg data and eye data produces an accuracy of about 85%. All the ASD participants were classified correctly. Whereas two out of the three Typically developing were classified correctly.The performance of this model is better than the one with only Standard Deviation data but same as that with only eye data.

- Deep Neural Network

With PCA
Num Components = 11

Loss: 0.80
Accuracy: 0.85

Predicted	ASD	Typical	All
Actual			
ASD	3	1	4
Typical	0	3	3
All	3	4	7

DNN with standard deviation EEG and eye data produces an accuracy of about 85%. Three out of the four ASD participants were classified correctly. Whereas all the Typically developing participants were classified correctly. The performance of this combined model is better than that of only eye data but the same as that with only SD EEG data.

Without PCA
Num Components = 159

Loss: 0.96
Accuracy: 0.57

Predicted	ASD	Typical	All
Actual			
ASD	2	2	4
Typical	1	2	3
All	3	4	7

Without PCA DNN's accuracy drops from 85% to 57%.

- <u>Gaussian Naive Bayes</u>

With PCA
Number of component = 24
Acc: 0.57

Predicted	ASD	Typical	All
Actual			
ASD	3	1	4
Typical	2	1	3
All	5	2	7

Gaussian Naive Bayes with Standard Deviation EEG data and eye data using PCA produces 57% accuracy. Three out of the four ASD participants were classified correctly. And only one out of the three typically developing participants were classified correctly. This combined model's performance is far lower than that with just eye data and only SD EEG data.

Without PCA
Number of component = 159
Acc: 0.71

Predicted	ASD	Typical	All
Actual			
ASD	4	0	4
Typical	2	1	3
All	6	1	7

Without PCA the accuracy shifts from 57% to 71%. All the ASD participants are classified correctly. Whereas only one out of the three Typically Developing participants are diagnosed correctly. The performance is far lower than that with just eye data and the same as the one with just SD EEG data.

C. Entropy with Eye Movement

- <u>SVM</u>

With and Without PCA
Num Components =24, 288

Loss: 19.73
Accuracy: 0.428

Predicted	ASD	Typical	All
Actual			
ASD	0	4	4
Typical	0	3	3
All	0	7	7

SVM with Entropy eeg data and eye data produces an accuracy of about 42%. All the ASD participants were misclassified. Whereas all the Typically developing participants were classified correctly. The performance of this model is worse than the one with just eye data and just entropy data without PCA.

- Logistic Regression

With and without PCA
Num Components = 24

Accuracy : 0.85

Predicted	ASD	Typical	All
Actual			
ASD	4	0	4
Typical	1	2	3
All	5	2	7

Logistic regression with entropy EEG and Eye data produces and accuracy of about 85%. All the ASD participants were classified correctly. Whereas two out of the three Typically developing participants were classified correctly. The performance of this combined model is the same as that with just eye data but better than the one with just entropy eeg data.

- <u>Deep Neural Network</u>

  With PCA
  Num Components = 24
  Loss: 0.7
  Accuracy: 0.714

Predicted	ASD	Typical	All
Actual			
ASD	2	2	4
Typical	0	3	3
All	2	5	7

DNN with Entropy EEG data and eye data produces an accuracy of about 71%. Two out of the four ASD participants were classified correctly. Whereas all the Typically developing participants were classified correctly. This combined model's performance is the same as the one with just eye data but better than that with just Entropy eeg data.

Without PCA

Loss: 0.726
Accuracy: 0.57

Predicted	ASD	Typical	All
Actual			
ASD	3	1	4
Typical	2	1	3
All	5	2	7

Without using PCA the accuracy drops from 71% to 57%. The comparison to the other models is the same as the model with PCA.

- <u>Gaussian Naive Bayes</u>

With PCA
Num Components = 24
Accuracy: 0.85

Predicted	ASD	Typical	All
Actual			
ASD	3	1	4
Typical	0	3	3
All	3	4	7

Gaussian Naive Bayes with Entropy EEG data and Eye data using PCA produces about 84% accuracy. Three out of the four ASD participants were classified correctly. Whereas all the Typically developing participants were classified correctly. The performance of this combined model is same as that with just eye data but better than the one with just entropy eeg data.

Without PCA
NC = 288

Accuracy: 1.0

Predicted	ASD	Typical	All
Actual			
ASD	4	0	4
Typical	0	3	3
All	4	3	7

Without PCA the Gaussian Naive Bayes model with combined entropy eeg and eye data produces a 100% accuracy. All the participants are classified correctly. The performance of this combined model is the same as the one with just eye data but better than the one with just entropy eeg data.

D. DFFT with Eye Movement

- <u>SVM</u>

  With and Without PCA
  Loss: 19.73
  Accuracy: 0.428

Predicted	ASD	Typical	All
Actual			
ASD	0	4	4
Typical	0	3	3
All	0	7	7

- <u>Logistic Regression</u>
  With and without PCA
  Accuracy: 0.85

Predicted	ASD	Typical	All
Actual			
ASD	3	1	4
Typical	0	3	3
All	3	7	7

- **DNN**
  Without PCA

  Loss: 0.84
  Accuracy: 0.42

Predicted	ASD	Typical	All
Actual			
ASD	3	1	4
Typical	3	0	3
All	6	1	7

With PCA

Loss: 0.74
Accuracy: 0.71

Predicted	ASD	Typical	All
Actual			
ASD	3	1	4
Typical	1	2	3
All	4	3	7

- **Gaussian Naive Bayes**

  With and Without PCA
  Accuracy: 0.714

Predicted	ASD	Typical	All
Actual			
ASD	4	0	4
Typical	2	1	3
All	6	1	7

8.6 Using Sequential Feature Selection Algorithm

8.6.1 Combination of EEG and Eye data

A) Gaussian Naive Bayes

- Mean

Accuracy: 1.0

Predicted	ASD	Typical	All
Actual			
ASD	4	0	4
Typical	0	3	3
All	4	3	7

- Standard Deviation

Accuracy: 1.0

Predicted	ASD	Typical	All
Actual			
ASD	4	0	4
Typical	0	3	3
All	4	3	7

- FFT
  Accuracy: 1.0

Predicted	ASD	Typical	All
Actual			
ASD	4	0	4
Typical	0	3	3
All	4	3	7

- Shannon Entropy

  Accuracy: 1.0

Predicted	ASD	Typical	All
Actual			
ASD	4	0	4
Typical	0	3	3
All	4	3	7

B) Logistic Regression

- Shannon Entropy

  Accuracy: 1.0

Predicted	ASD	Typical	All
Actual			
ASD	4	0	4
Typical	0	3	3
All	4	3	7

- FFT
  Accuracy: 0.857

Predicted	ASD	Typical	All
Actual			
ASD	4	0	4
Typical	1	2	3
All	5	2	7

- Mean

  Accuracy: 1.0

Predicted	ASD	Typical	All
Actual			
ASD	4	0	4
Typical	0	3	3
All	4	3	7

- Standard Deviation

  Accuracy: 1.0

Predicted	ASD	Typical	All
Actual			
ASD	4	0	4
Typical	0	3	3
All	4	3	7

C) SVM

- Shannon Entropy

  Accuracy: 1.0

Predicted	ASD	Typical	All
Actual			
ASD	4	0	4
Typical	0	3	3
All	4	3	7

- FFT

Accuracy: 0.42

Predicted	ASD	Typical	All
Actual			
ASD	2	2	4
Typical	2	1	3
All	4	3	7

- Mean

Accuracy: 0.42

Predicted	ASD	Typical	All
Actual			
ASD	0	4	4
Typical	0	3	3
All	0	3	7

- Standard Deviation

Accuracy: 0.71

Predicted	ASD	Typical	All
Actual			
ASD	2	2	4
Typical	0	3	3
All	2	5	7

D) DNN

- Mean

Accuracy: 0.42

Predicted	ASD	Typical	All
Actual			
ASD	3	1	4
Typical	3	0	3
All	6	1	7

- Standard Deviation

Accuracy: 0.57

Predicted	ASD	Typical	Al l
Actual			
ASD	2	2	4
Typical	1	2	3
All	3	4	7

- Entropy
  Accuracy: 0.57

Predicted	ASD	Typical	All
Actual			
ASD	3	1	4
Typical	2	1	3
All	5	2	7

- FFT

  Accuracy: 0.42

Predicted	ASD	Typical	All
Actual			
ASD	2	2	4
Typical	2	1	3
All	4	3	7

8.5.2 EEG data

A) Logistic

- Mean

  Accuracy: 0.57

Predicted	ASD	Typical	All
Actual			
ASD	1	3	4
Typical	0	3	3
All	1	6	7

- Standard Deviation

  Accuracy: 0.57

Predicted	ASD	Typical	All
Actual			
ASD	2	2	4
Typical	1	2	3
All	3	4	7

- FFT

  Accuracy: 0.42

Predicted	ASD	Typical	All
Actual			
ASD	1	3	4
Typical	1	2	3
All	2	5	7

- Entropy

Accuracy: 0.428

Predicted	ASD	Typical	All
Actual			
ASD	0	4	4
Typical	0	3	3
All	0	5	7

C) SVM

- Standard Deviation

Accuracy: 0.428

Predicted	ASD	Typical	All
Actual			
ASD	1	3	4
Typical	1	2	3
All	2	5	7

- Mean
  Accuracy: 0.428

Predicted	ASD	Typical	All
Actual			
ASD	0	4	4
Typical	0	3	3
All	0	7	7

- FFT

  Accuracy: 0.42

Predicted	ASD	Typical	All
Actual			
ASD	1	3	4
Typical	1	2	3
All	2	5	7

- Entropy

Accuracy: 0.428

Predicted	ASD	Typical	All
Actual			
ASD	0	4	4
Typical	0	3	3
All	0	7	7

D) GNB

- Standard Deviation
Accuracy: 0.428

Predicted	ASD	Typical	All
Actual			
ASD	2	2	4
Typical	2	1	3
All	4	3	7

- Mean

Accuracy: 0.57

Predicted	ASD	Typical	All
Actual			
ASD	1	3	4
Typical	0	3	3
All	1	6	7

- FFT

Accuracy: 0.42

Predicted	ASD	Typical	All
Actual			
ASD	3	1	4
Typical	3	0	3
All	6	1	7

- Entropy

Accuracy: 0.285

Predicted	ASD	Typical	All
Actual			
ASD	1	3	4
Typical	2	1	3
All	3	4	7

E) DNN

- Standard Deviation

Accuracy: 0.42

Predicted	ASD	Typical	All
Actual			
ASD	0	4	4
Typical	0	3	3
All	0	7	7

- Mean

Accuracy: 0.57

Predicted	ASD	Typical	All
Actual			
ASD	1	3	4
Typical	0	3	3
All	1	6	7

- FFT

Accuracy: 0.71

Predicted	ASD	Typical	All
Actual			
ASD	3	1	4
Typical	1	2	3
All	4	3	7

- Entropy

Accuracy: 0.42

Predicted	ASD	Typical	All
Actual			
ASD	2	2	4
Typical	2	1	3
All	4	3	7

8.5.3 Eye Data

- SVM

Accuracy: 1.0

Predicted	ASD	Typical	All
Actual			
ASD	4	0	4
Typical	0	3	3
All	4	3	7

- Logistic

Accuracy: 1.0

Predicted	ASD	Typical	All
Actual			
ASD	4	0	4
Typical	0	2	3
All	4	3	7

- Gaussian Naive Bayes

Accuracy: 1.0

Predicted	ASD	Typical	All
Actual			
ASD	4	0	4
Typical	0	3	3
All	4	3	7

- DNN

Accuracy: 1.0

Predicted	ASD	Typical	All
Actual			
ASD	4	0	4
Typical	0	3	3
All	4	3	7

9 COMPARISON

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.

9.1 Training and Test

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.

9.1.1 Without Using Sequential Feature Selection Algorithm

The following table is without using Sequential Feature Selection algorithm.

Table 1

Eye Vs SD EEG Vs Combination

	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

Eye Vs Mean EEG Vs Combination

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

The above table 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

Eye Vs FFT EEG Vs Combination

	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 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

Eye Vs Entropy EEG Vs Combination

	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.

9.1.2 Using Sequential Feature Selection Algorithm

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

Table 5

Eye Vs Entropy EEG Vs 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%

Table 6

Eye Vs EEG FFT Vs 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%

Table 7

Eye Vs EEG Standard Deviation Vs 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%

Table 8

Eye Vs EEG mean Vs 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%

9.2 10 x 2 Cross Validation

10 x 2 Cross Validation repeated 200 times shows more confirming results than 5x 2 Cross Validation.

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

Table 9

10 x 2 Cross validation with Sequential Feature Selection

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

Table for Combined Features using Sequential Feature Selection

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%

Below is the comparison bar graph of FFT EEG and combined FFT EEG using Sequential Feature Selection.

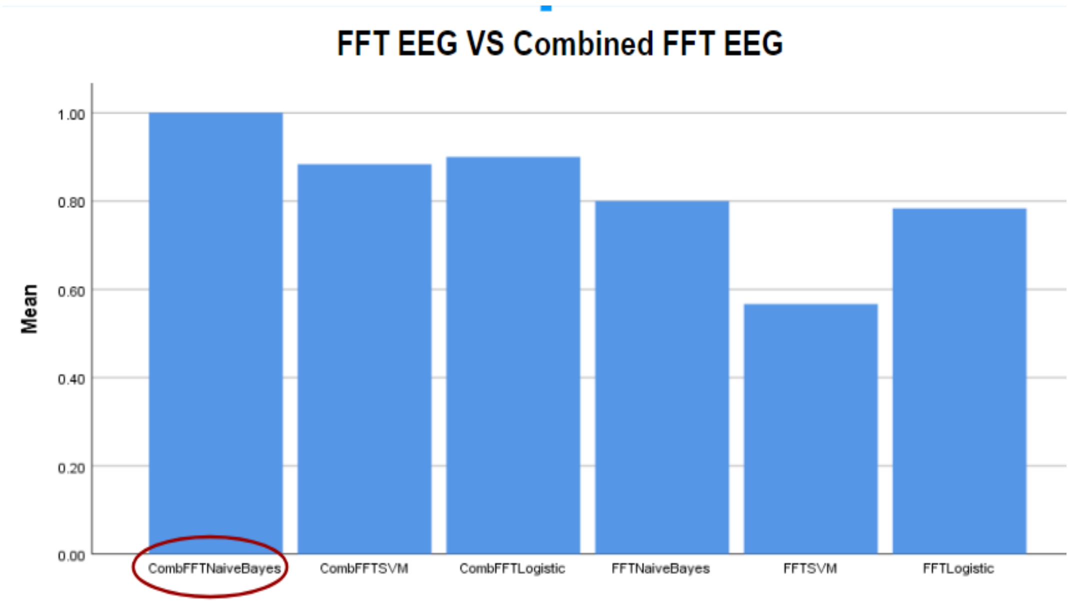

Fig 23. FFT EEG Vs Combined FFT EEG

In the above graph 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. Below is the image that shows the comparison between Entropy EEG models and Combined Eye and Entropy EEG models.

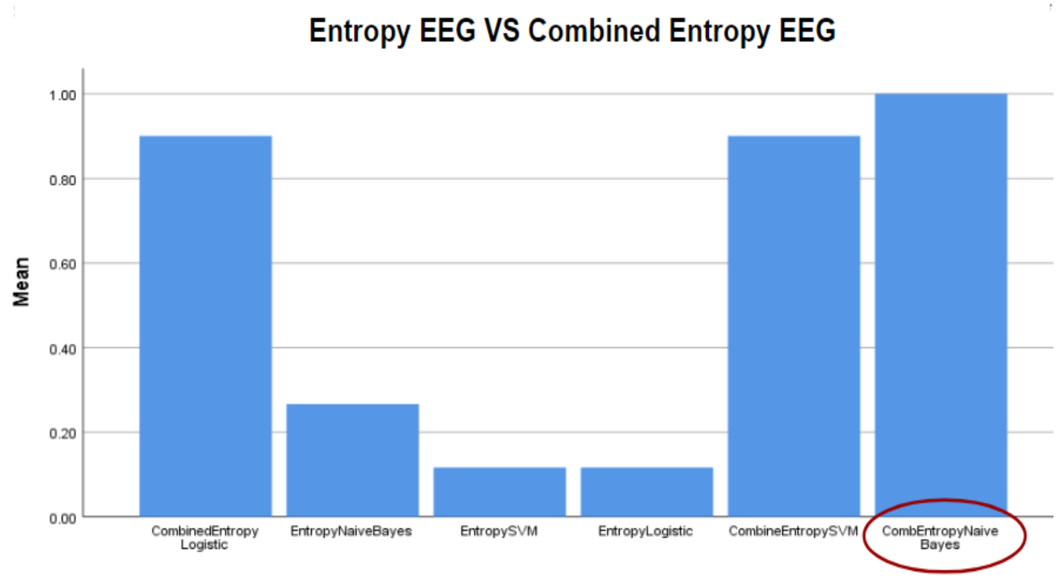

Fig 24. Entropy EEG Vs Combined Entropy EEG

In the above graph 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.

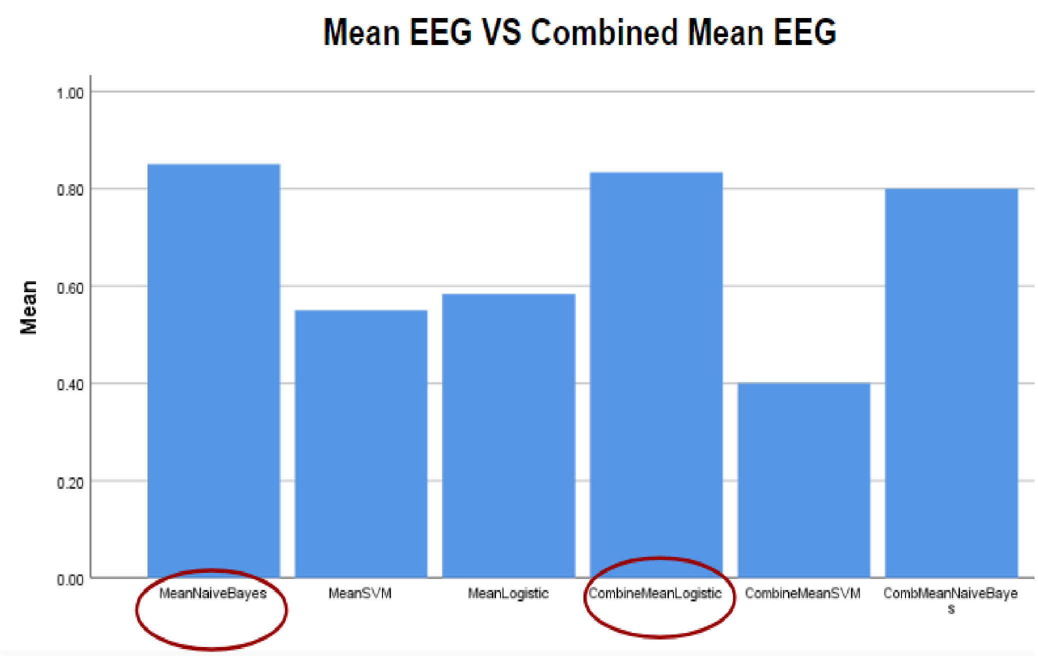

Fig 25. Mean EEG Vs Combined Mean EEG

In the above graph we can see than highest performing models with around 82% average accuracy are Naive Bayes with just Mean EEG and Logistic with combined data.

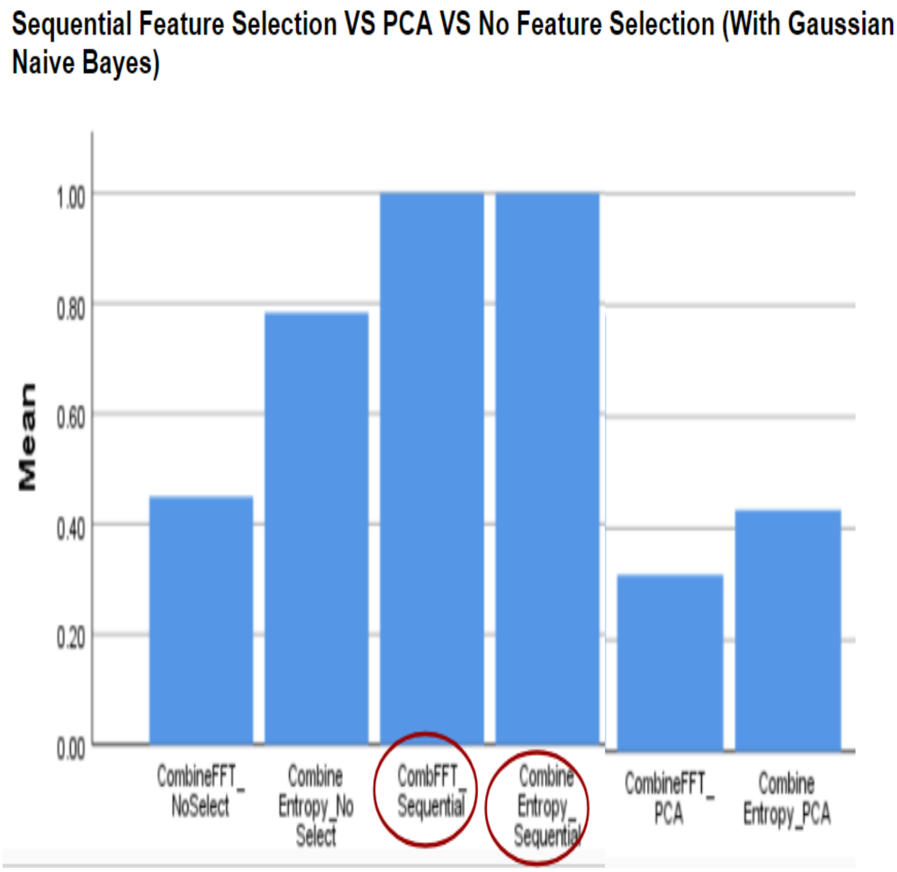

Fig 26. Feature Selection Algorithm Comparison

In the above graph we can see than Sequential Feature Selection algorithm performs the best.

10 CONCLUSION

In this research I have made an analysis and comparison between EEG, Eye and combined data. I 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. I found that for some models with PCA did better while for some without PCA did better. Like for 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.

11 FUTURE WORK

In the future there are a lots of areas of improvement and a lot more comparison can be made. As this is a preliminary work towards finding the optimal types of features for ASD diagnosis. The main potential work are as follows:

1) Using various Feature Selection Method

   In this thesis I 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.

2) Extracting more EEG features

   Like in this thesis I compared Mean, Average, Standard Deviation, DFFT and Shannon Entropy of EEG. There are a lots of other features that could be extracted from EEG like RR, DET, ENTR, energies, and AR coefficients of wavelets etc.

3) Using more machine learning models

   I have only used four different machine learning models; SVM, Logistic Regression, DNN, and Naive Bayes. But there are many other machine learning models that could be used like KNN, Decision Trees, Random Forests, CNN, and RNN etc. And one of them could be even better than the Naive Bayesian models.

4) Increasing the data size

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 33 participants is too low to confirm our results. However, this is a first step towards developing an optimal Autism Diagnosis system.

REFERENCES

[1] Grossi, E., Olivieri, C., & Buscema, M. (2017). Diagnosis of autism through EEG processed by advanced computational algorithms: A pilot study. *Computer Methods and Programs in Biomedicine,142*, 73-79. doi:10.1016/j.cmpb.2017.02.002

[2] 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

[3] Jamal, W., Das, S., Oprescu, I., Maharatna, K., Apicella, F., & Sicca, F. (2014). Classification of autism spectrum disorder using supervised learning of brain connectivity measures extracted from synchrostates. *Journal of Neural Engineering,11*(4), 046019.

[4] 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.

[5] A. Dejman, A. Khadem and A. Khorrami, "Exploring the disorders of brain effective connectivity network in ASD: A case study using EEG, transfer entropy, and graph theory," *2017 Iranian Conference on Electrical Engineering (ICEE)*, Tehran, 2017, pp. 8-13.

[6] D. Begum, K. M. Ravikumar and K. N. Vykuntaraju, "An initiative to classify different neurological disorder in children using multichannel EEG signals," *2016*

IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, 2016, pp. 1563-1566.

[7] W. Khazaal Shams and A. W. Abdul Rahman, "Characterizing autistic disorder based on Principle Component Analysis," *2011 IEEE Symposium on Industrial Electronics and Applications*, Langkawi, 2011, pp. 653-657.

[8] Djemal, Ridha et al. “EEG-Based Computer Aided Diagnosis of Autism Spectrum Disorder Using Wavelet, Entropy, and ANN.” *BioMed Research International* 2017 (2017): 9816591. *PMC*. Web. 18 Apr. 2018.

[9] Raja, Laxmi, and Mohana Priya. “Neural Network based Classification of EEG signals for diagnosis of Autism Spectrum Disorder.” *International Journal of Pharma and Bio Sciences*, Apr. 2017, www.ijpbs.net/cms/php/upload/5952_pdf.pdf.

[10] S. Canavan *et al*., "Combining gaze and demographic feature descriptors for autism classification," *2017 IEEE International Conference on Image Processing (ICIP)*, Beijing, 2017, pp. 3750-3754.

[11] I. Torii, K. Ohtani and N. Ishii, "Measurement of Ocular Movement Abnormality in Pursuit Eye Movement (PEM) of Autism Spectrum Children with Disability," *2016 4th Intl Conf on Applied Computing and Information Technology/3rd Intl Conf on Computational Science/Intelligence and Applied Informatics/1st Intl Conf on Big Data, Cloud Computing, Data Science & Engineering (ACIT-CSII-BCD)*, Las Vegas, NV, 2016, pp. 235-240.

[12] 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.

[13] T. Vu *et al*., "Effective and efficient visual stimuli design for quantitative autism screening: An exploratory study," *2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, Orlando, FL, 2017, pp. 297-300.

[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

[19] https://ieeexplore.ieee.org/xpls/icp.jsp?arnumber=6498036

[20]
https://ieeexplore.ieee.org/document/4338305/

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

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