

EEG Channel Selection for AR Model Based ADHD Classification

Juan L. Lopez Marcano[§], Martha Ann Bell[†], and A. A. (Louis) Beex[§]

[§]DSPRL – Wireless@VT – ECE Department, [†]Psychology Department
Virginia Tech, Blacksburg VA, USA

Abstract— As of today, diagnosis of ADHD is highly dependent on subjective observations, which has motivated researchers to investigate quantitative methods for the discrimination of ADHD and Non-ADHD subjects using EEG data. The goal of the effort reported here is to classify subjects with high accuracy, as well as to do so based on a select few channels. By making use of AR model features, several classifiers were found to achieve high performance; accuracy above 90% for a K Nearest Neighbor classifier and Area Under the Curve over 0.98 at Equal Error Rate below 0.05 for a Gaussian Mixture Model-Uniform Background Model classifier based on combinations of as few as 2 and 3 EEG channels.

Key words: ADHD, classification, channel reduction

I. INTRODUCTION

According to the Diagnostic and Statistical Manual of Mental Disorders (DSM-V) [1], attention deficit hyperactivity disorder (ADHD) is a condition characterized by high levels of inattention, hyperactivity, and impulsivity. It is estimated that 11% of children ages 4 to 17 in the US are affected by ADHD [2]. Diagnosis of ADHD is done by comparing observed symptoms with the symptoms described in the DSM-IV. While the latter recognizes three subtypes of ADHD, the labeling of the data available for this work allows differentiation between ADHD and Non-ADHD only.

As of today, diagnosis of ADHD is highly dependent on observation of symptoms by parents, behavioral scientists, and physicians, which is subjective. As a result, finding quantitative techniques to aid in the diagnosis of ADHD has gained attention. These techniques include estimating power in frequency bands, phase-synchrony, coherence, and supervised and semi-supervised learning [3, 4, 5, 6, 7, 8]. The results obtained in these studies have shown that ADHD (A) and Non-ADHD (NA) subjects are, to some extent, separable in several feature domains.

This study concerns the selection and reduction in the number of EEG channels that is adequate to classify NA and A subjects at a desired level of performance. Motivated by encouraging previous results [6], the task at hand is to explore whether there is a better combination of fewer than 5 channels that will further maximize discrimination of A and NA subjects while also reducing computational cost. The latter is a consideration towards eventual development of an efficient portable prototype for point of care diagnostic purposes. KNN (k-nearest neighbor) classifiers and GMM-UBM models (Gaussian Mixture Models – Universal Background Model) were used for classification. The parameters of auto-regressive (AR) models are used as

features. Datasets from 4 subjects (2 NA and 2 A) were used for training, while datasets from 4 other subjects (1 NA and 3 A) were used for testing. The performance associated with each selection of channels was analyzed in terms of classification accuracy (True Positive + True Negative divided by the number of tests) for KNN, and AUC (area under the curve) and EER (equal error rate) for GMM-UBM.

The structure of the paper is as follows: Section II provides an overview of how EEG has been used for discrimination between ADHD and Non-ADHD subjects. In Section III the methods used are described. Section IV covers the experiments made and the results. Finally, the conclusion is given in Section V.

II. RELATED WORK

In 1999, a study reported that the θ/β power ratio of ADHD subjects – during resting conditions – was higher than that of control subjects [7]. In said study, the power in frequency bands was obtained by computing PSD estimates from the FFT. To test the hypothesis, the θ/β power ratio for all the control subjects was averaged, and the decision made to classify as ADHD subjects those whose θ/β power ratio was 1.5 standard deviations above the average θ/β power ratio of control subjects. The sensitivity reported was 87% and specificity 94%. Numerous studies have been performed to validate the results. Since then, when attempting to replicate the method [7], the accuracy (sensitivity + specificity divided by 2) that researchers have found varies from 50 to 94% [8, 9]. These results suggest that signal power is not enough to produce a diagnosis, but could perhaps be used to pre-screen subjects. Teachers and parents reportedly [10] can identify ADHD with an accuracy ranging from 47% to 58%, i.e. slightly above chance.

Another study [3] used power in frequency bands along with semi-supervised learning in order to diagnose ADHD subjects. In this study, the power and power ratios in the α , β , θ , and γ frequency bands were computed and the mutual information criterion used to choose the least redundant features for training of a Gaussian support vector machine. EEG recordings of 10 subjects were used, and the accuracy of classification was 97%; the miss rate was not reported.

In our earlier publication [6], AR parameters, extracted from EEG during attention activity, and supervised learning were used for the classification of ADHD and Non-ADHD subjects based on a KNN classifier. AR(7) models were computed from windows of 2 s intervals, and a KNN classification accuracy between 85% and 95% was obtained. In addition, a confidence metric was derived from the vote count of the KNN classifier, in terms of the fraction of K nearest neighbors siding with the majority decision; the confidence metric ranged from 91% to 99%.

The effectiveness of event-related potentials (ERPs) has also been studied [4]; 74 control and 74 ADHD subjects performed a visual two-stimulus GO/NOGO task while their EEG data was recorded. Independent component analysis (ICA) performed on the ERPs, and using these features to train a SVM classifier, achieved 92% accuracy of classification (90% sensitivity and 94% specificity).

III. METHODS

An overview is provided in Section III.A of how the data were collected, in Section III.B of the features that were extracted, and in Section III.C of the classification techniques used for channel selection.

A. Data Collection

Children between the ages of 6 and 8 years visited the research lab as part of an ongoing longitudinal study (study procedures were approved by the Virginia Tech Institutional Review Board) focused on frontal lobe development from infancy through childhood. Information regarding diagnosis of ADHD was obtained via maternal report. EEG was recorded using a stretch cap (Electro-Cap, Inc Eaton, OH: E1-series cap) in the extended 10/20 system pattern. Recordings were made from 26 electrodes located equidistant across the scalp.

Electrode impedances were kept under 20k ohms. The electrical activity from each lead was amplified using separate bioamps (James Long Company, Caroga Lake, NY). During data collection, the high-pass filter was a single pole RC filter with a 0.1 Hz cut-off (3 dB or half-power point) and 6 dB/octave roll-off. The low-pass filter was a two-pole Butterworth type with a 100-Hz cut-off (3 dB or half-power point) and 12 dB/octave roll-off. The EEG signal was digitized at 512 samples per second for each channel so that data were not affected by aliasing. The acquisition software was Snapshot-Snapstream (HEM Data Corp, Southfield MI). Prior to the recording of each subject, a 10 Hz, 50 uV peak-to-peak sine wave was input through each amplifier and digitized for 30 sec. This signal was analyzed and the resulting power values used to calibrate the EEGs.

After the EEG electrodes were applied, children participated in eyes open, eyes closed, and quiet video baseline events to collect resting EEG data. Then the children completed a battery of cognitive tasks designed to assess various aspects of attention [11] using the child version [12] of the Attention Network Task (ANT) and various aspects of cognition associated with executive functions (e.g., number Stroop, Dimensional Change Card Sort Task, Digit Span Task).

In contrast with the θ/β power ratio approach, which used data during resting conditions, here data collected during the ANT were used in all subsequent analyses.

The ANT was designed to assess Posner’s brain-based attention networks [11] and yields measures of conflict, alerting, and orienting. The test requires the child to respond to a central target (a yellow fish on a light blue background) displayed on a computer screen and indicate whether the fish is facing left or right. The child is instructed to look at the fixation point, above or below which the target will appear. The target may appear with or without flankers (other fish),

which may or may not be congruent with respect to the direction they are facing. Reaction time responses to the alert cues, spatial cues, and flankers are manipulated to provide an assessment of the efficiency of each of the attention networks. The ANT is divided into 3 blocks of ~5 minutes each, with a brief rest period between blocks. The EEG during the first ANT block was used in the subsequent analyses reported here.

B. Feature Extraction

The features used are the autoregressive (AR) parameters a_k , extracted from finite length observation records (2 seconds) for the selected channels. The data is modeled as

$$x_n = -\sum_{k=1}^p a_k x_{n-k} + \varepsilon_n \quad (1)$$

where p is the order of the AR process and ε_n is the prediction error process. The AR model gives a compact representation (a small number of model parameters versus a large number of samples) for processes that are resonant, narrowband, or pseudo-sinusoidal, such as EEG rhythms; hence its proposed use here to generate features for ADHD detection. The Burg method was used for AR parameter estimation, after finding a reasonable order for the model using Akaike Information Criterion:

$$AIC(p) = N \ln(\lambda^2) + 2p \quad (2)$$

where N is the number of observed samples, and λ^2 is the estimated prediction error variance. The “best” estimate for the AR model order is the one minimizing $AIC(p)$.

C. Channel Selection

To find the best combination of just a few channels, two different classification methods were studied. These methods consist of iteratively using KNN classifiers and GMM-UBMs respectively to find the set of channels that suits each classification scheme best, in terms of performance.

KNN

KNN classifiers were trained and tested to rank the best combinations of channels based on the accuracy of classification achieved with them. For training, feature vectors extracted from 4 subjects (2 A and 2 NA) were used, and for testing, feature vectors extracted from 3 A and 1 NA subjects that were not part of the training dataset were used; all 30 possible unique combinations were used.

Ranking was done as follows: The EEG data from all possible 2-channel combinations was used for feature extraction. KNN classifiers were trained and tested, as explained in the previous paragraph, and the combination that had the highest mean classification accuracy was chosen. To find a 3rd channel, all possible 3-channel combinations that include the best 2 were used in feature extraction and then used for training and classification. This process was repeated to find the fourth channel.

GMM-UBM

A Gaussian Mixture Model (GMM) is a model for a probability density function (pdf) expressed as a weighted sum of Gaussian probability density functions [13]. The

main reason for using GMMs for classification problems is that Gaussian mixture densities can approximate any arbitrary pdf [13]. The pdf of a GMM λ is expressed as:

$$p(\mathbf{v} | \lambda) = \sum_{m=1}^M w_m g_m(\mathbf{v} | \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m) \quad (3)$$

where \mathbf{v} is an N -dimensional feature vector, w_m are the weights (non-negative and summing to 1), and g_m are the individual N -variate Gaussian pdfs, which have the following form:

$$g_m(\mathbf{v} | \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m) = \frac{1}{(2\pi)^{N/2} |\boldsymbol{\Sigma}_m|} \exp\left(-\frac{1}{2} \|\mathbf{v} - \boldsymbol{\mu}_m\|_{\boldsymbol{\Sigma}_m^{-1}}^2\right) \quad (4)$$

where $\boldsymbol{\mu}_m$ is an N -dimensional column vector and $\boldsymbol{\Sigma}_m$ is an $N \times N$ covariance matrix. The parameters $\boldsymbol{\mu}_m$, w_m , and $\boldsymbol{\Sigma}_m$ identify the M mixture components and weights of the GMM. In this study, M was set to 4 because training was done using 4 subjects.

To train the GMMs, i.e. finding the model parameters, the expectation maximization (EM) algorithm was used.

In analogy with the speaker verification approach [14], UBMs were found using various combinations of features extracted from Non-ADHD subjects (impostors). Models were also found to fit the class of ADHD subjects (targets). For classification, the log-likelihood ratio (LLR) is used, i.e. the ratio of the likelihood of a test vector \mathbf{v}_t belonging to the ADHD model over the likelihood of \mathbf{v}_t belonging to the universal background model. If the LLR is greater than or equal to zero, the subject is classified as ADHD, otherwise the subject is classified as Non-ADHD.

$$LLR = \log\left(\frac{p(\mathbf{v}_t | \lambda_{ADHD})}{p(\mathbf{v}_t | \lambda_{UBM})}\right) \quad (5)$$

For training, feature vectors extracted from 4 subjects (2 A and 2 NA) were used, and for testing, feature vectors extracted from 3 A subjects and 1 NA subject that were not part of the training dataset were used. Channel ranking for GMM-UBMs was done in the same fashion as described for KNN classifiers.

Figure 1 summarizes the training and testing process, for both the KNN and GMM-UBM approaches.

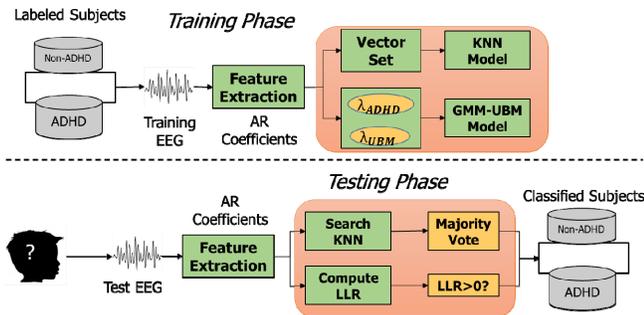


Fig. 1: Training and testing diagram

IV. EXPERIMENTS

Consistent with the previously reported approach [6], AR models of order 7 were computed in order to compensate for the tendency of AIC to overestimate the order of AR models [15]. For every test channel, AR(7) parameters were computed, and feature vectors were formed as the concatenation of the AR coefficients of all test channels. Thus, when searching for the best combination of 2 channels, 14-D feature vectors were obtained; when searching for the best third channel to add to the best 2-channel combination, 21-D feature vectors were used; when searching for a fourth channel to add to the latter combination, 28-D feature vectors were used. Lastly, in our earlier publication [6], 35-D feature vectors were used.

By using 4 subjects for training (2 A and 2 NA) and 4 others (3 A and 1 NA) for testing, the accuracy of classification was explored across all 30 unique permutations given the data available (5 A and 3 NA). For training and testing, AR models were computed from windows of 2 s duration, using an overlap of 50%, during the duration of ANT. Note that this approach produces many (in the range of 240 to 260) test vectors for a single test subject and for each test vector a decision is made so that a distribution of decisions results. For KNN, the best classification performance resulted from using 51 nearest neighbors.

Accuracy, when KNN classifiers are used, is defined as the number of true positives (TP) plus the number of true negatives (TN) over the total number of tests:

$$Accuracy = \frac{TP + TN}{\#tests} \quad (6)$$

For GMM-UBMs, the performance metric is compactly defined in terms of ROC (receiver operating characteristic) curve derived values, such as AUC and EER.

A. KNN

For this channel ranking scheme, the accuracy of classification using all possible and unique 2-channel combinations was investigated first. Once the combination that lead to the highest accuracy was found, the investigation focused next on which channel(s) could be added thereto to obtain the best 3 and 4-channel combinations.

Figure 2 shows the accuracy of classification for all 231 ($=21 \times 22/2$) possible unique 2-channel combinations zoomed in to the range from 0.8 to 0.93 with color. As a result, in Fig. 1 many entries in the matrix are blue or dark blue, meaning that the accuracy of classification was below 0.86. In this experiment, the best classification performance was realized when pair Fc1-Pz was used, which achieved a classification accuracy CDF (cumulative distribution function) characterization of $\{0.8933 - 0.9312 - 0.9637\}$, which are the 5th percentile, the mean, and the 95th percentile respectively. The latter percentiles provide an idea of how concentrated the accuracy (viewed as a random variable) is about the mean. The Fc1-Pz pair is followed by Fc1-Cp2, with classification performance of $\{0.8646 - 0.9238 - 0.9645\}$. Other combinations, such as Cp2-Pz, Fc1-Fz, and

Fc2-Pz ranked high as well (third, fourth, and fifth respectively), but achieved accuracies of less than 0.92.

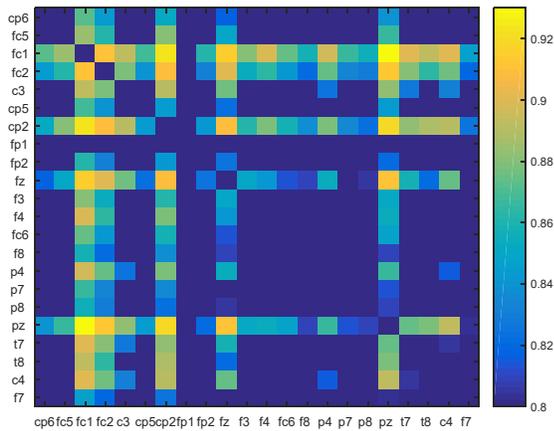


Fig. 2: Accuracy of 2-channel combinations in matrix form

Since Fc1 and Pz yielded the highest accuracy of classification, the Fc1-Pz pair was used to then search for the best 3-channel combination.

Figure 3 shows the accuracies achieved with all 3-channel combinations that include Fc1-Pz.

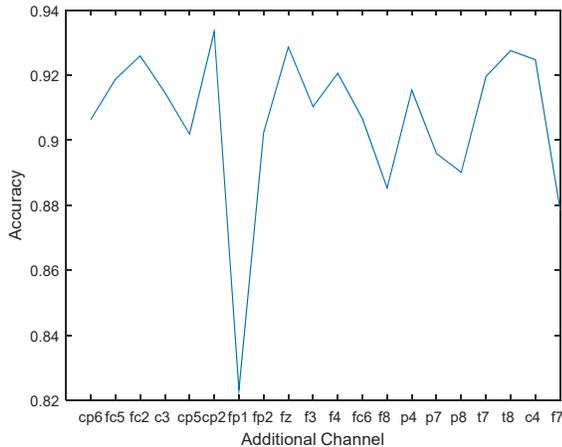


Fig. 3: Accuracy of 3-channel combinations that include Fc1-Pz

As seen, the 3-channel combination that yields the highest accuracy is Fc1-Pz-Cp2, for which the CDF values of classification performance are $\{0.8988 - 0.9334 - 0.9690\}$.

The combination Fc1-Pz-Fz yields the second best performance results, with accuracy of performance CDF values of $\{0.8831 - 0.9286 - 0.9581\}$. Figure 2 shows interesting results because Fc1-Cp2 ranked second in the 2-channel experiment, and Pz, Fz, and Fc2 show up in the third, fourth, and fifth best 2-channel combinations. It is also worth noting that mean accuracy does not increase by much when adding Cp2: an increase from 0.9312 to 0.9334.

Continuing along the lines above, the best 4-channel combination that includes Fc1-Pz-Cp2 is found. The combination Fc1-Pz-Cp2-T8 yields classification CDF values of $\{0.8996 - 0.9331 - 0.9775\}$. Note that while Fc1-

Pz-Cp2-T8 is the best 4-channel combination, accuracy is about the same (slightly higher 5th percentile, but slightly lower mean) as for its 3-channel counterpart, which suggests that a 3-channel combination would be more favorable in the feature domain used in this study. However, if computational complexity is to be minimized, the Fc1-Pz combination should be considered, given that adding Cp2 will only add 0.22% of accuracy.

B. GMM-UBM

In this section, the best 2, 3, and 4-channel combinations for GMM-UBM are investigated. The procedure followed is similar to that for KNN, except that the best combinations are chosen in terms of AUC and EER performance.

Figure 4 shows the performance for all 231 unique 2-channel combinations. Figure 4 is a matrix where the values above the diagonal represent the AUC and the values below the diagonal represent 1-EER (Complementary EER). To find the best combinations, AUC should be maximized and EER should be minimized, meaning that $1 - \text{EER}$ should be maximized as well.

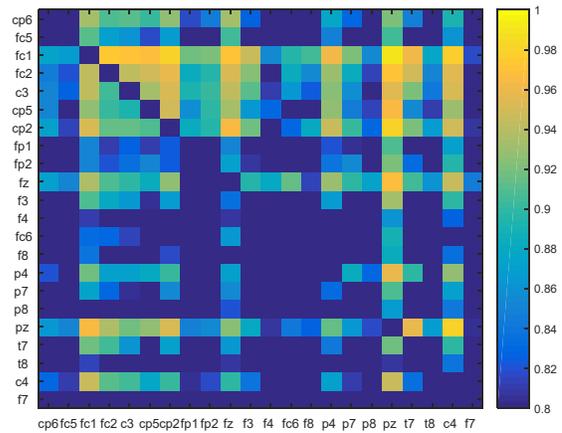


Fig. 4: AUC (above diagonal) and 1-EER (below diagonal) for all 2-channel combinations

In Fig. 4, many elements are blue and dark blue because the image was zoomed to the range from 0.8 to 1. Just as for Fig. 1, Fc1-Pz provides the best performance: Mean AUC of 0.9899 and mean EER of 0.0357. The 95th and 5th percentiles of the AUC of the Fc1-Pz combination are 0.9996 and 0.9730 respectively, and the 95th and 5th percentile of EER for this combination are 0.0700 and 0.0091 respectively. The 95th percentile of the AUC indicates that only 5% of the AUC exceed the 95th percentile, and the 5th percentile indicates that 95% of all the AUC exceed the 5th percentile. Similarly, for the EER, 5% of all EER exceed the 95th percentile and 95% of all EER exceed the 5th percentile. Interestingly enough, the second best 2-channel combination appears to be Cp2-Pz, which was also highly ranked with KNN. Cp2-Pz achieves AUC CDF values of $\{0.9577 - 0.9806 - 0.9982\}$ and EER CDF values of $\{0.0109 - 0.0491 - 0.0882\}$. Pz-C4 achieves a slightly higher AUC, but at the expense of EER, with CDF values of 0.0182 - 0.0536 - 0.1222}. Combinations involving Fc1, Fc2, Pz, and Cp2 are in the top 10 2-channel combinations, but their AUC are below 0.9800 and their EER above 0.0500.

Figure 5 shows the effect of adding one channel to Fc1-Pz. In Fig. 5, the AUC means fluctuate between 0.9700 and 0.9900. Similarly, the EER means fall between 0.0300 and 0.0900. The best 3-channel combination is Fc1-Pz-Cp2, with AUC CDF values of {0.9643 - 0.9867 - 0.9990} and EER CDF values of {0.0086 - 0.0367 - 0.0700}.

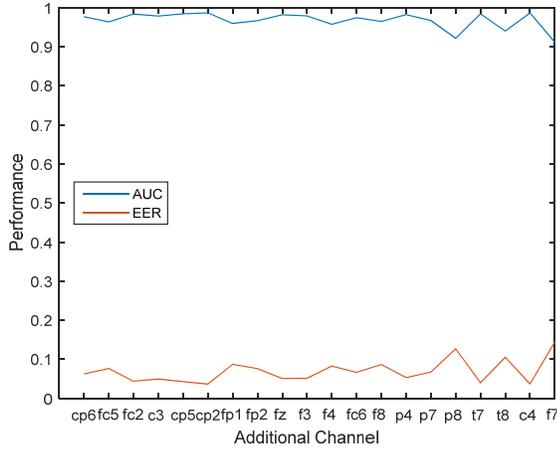


Fig. 5: AUCs and EERs of all 3-channel combinations that include Fc1-Pz

The second best 3-channel combination is Fc1-Pz-C4, with AUC CDF values of {0.9622 - 0.9865 - 0.9992} and EER CDF values of {0.0109 - 0.0373 - 0.0745}. The EER means for all other combinations exceed 0.0400 and the corresponding AUC means are below 0.9850. Note that most 3-channel combinations involving Fc1 and Pz achieve high AUC and low EER. As shown in Fig. 5, it is not advisable to combine Fc1-Pz with T8, P8, or F7.

This experiment suggests that a 2-channel combination may achieve higher performance than 3-channel combinations when using GMM-UBM and AR models. The mean AUC and mean EER of Fc1-Pz-Cp2 are 0.9867 and 0.0367 respectively. On the other hand, the mean AUC and mean EER of Fc1-Pz are 0.9899 and 0.0357, which indicates that on average performance deteriorates when a third channel is added to Fc1-Pz. In addition, the AUC and EER of Fc1-Pz-Cp2 are more spread than those of Fc1-Pz. Since feature extraction and classification are less computationally demanding for the 2-channel combinations, which also happens to improve performance for the feature set and classification scheme of this section, leads to the strong suggestion that Fc1-Pz be used.

For completeness, the effect of adding a channel to Fc1-Pz-Cp2 was also explored. The best 4-channel combination, Fc1-Pz-Cp2-T7, has a mean AUC of 0.9844 and a mean EER of 0.0367. For this combination, the 95th and 5th percentiles of the AUCs are 0.9988 and 0.9546 respectively, and the 95th and 5th percentiles of the EERs are 0.0705 and 0.0159. Although the mean EER did not change and the AUC is slightly smaller than that for the Fc1-Pz-Cp2 combination, computational complexity is not favorable. When using Fc1-Pz, feature vectors are 14-D, whereas they are 28-D when using Fc1-Pz-Cp2-T7. Besides, mean AUC is 0.9899 when using Fc1-Pz only. As a result, this experiment suggests that the number of channels be kept at 2.

Figure 6 illustrates a common ROC obtained when training and testing GMM-UBMs using pair Fc1-Pz.

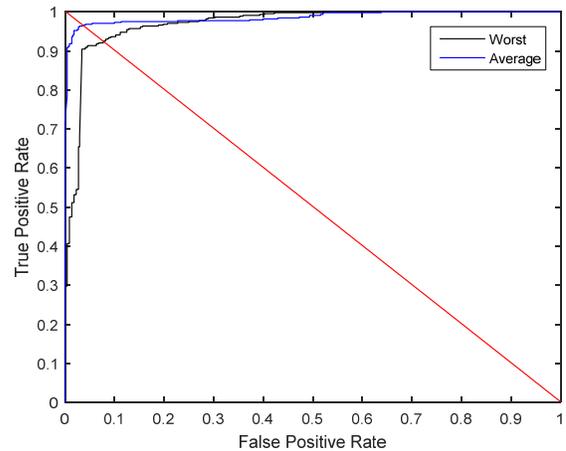


Fig. 6: ROC plots when pair Fc1-Pz is used in GMM-UBM

The closer the ROC curve comes to approximating the top-left corner the better the performance. Of all 30 test case permutations, each producing its own ROC, the worst case (black) and the averaged (blue) ROC are shown.

The area under the ROC curve is the AUC, and the intersection between the ROC curve and the red line, which is when the false acceptance (FP) and false rejection (FN) rates are equal, is the EER. The AUC for the average ROC shown is 0.9860 and its EER is 0.0364. What may be even more meaningful in practice is that for the worst case ROC, AUC is 0.9805 and EER is 0.0700, meaning that for all 30 test case configurations AUC is higher than 0.9805 and EER is 7% or less. Each of these ROC curves is based on using between 240 and 260 test vectors from each test subject, i.e. incorporating results from 1000 to 1100 test vectors.

An alternative way of presenting EER in a more detailed fashion is by way of detection error tradeoff (DET) curves, as shown in Fig. 7, for the Fc1-Pz pair using GMM-UBM.

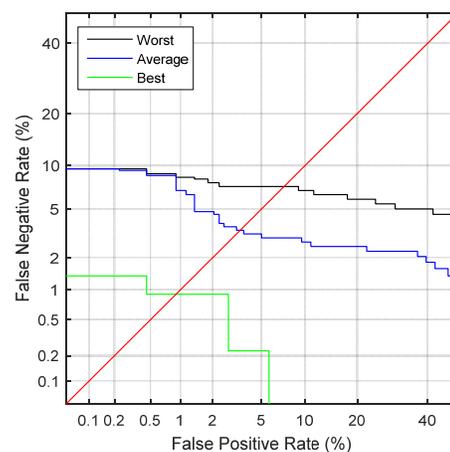


Fig. 7: DET curves when pair Fc1-Pz used in GMM-UBM

In a DET curve false negative rate (FNR) is plotted versus false positive rate (FPR) on logarithmic axes. Figure 7

shows the best, average, and worst of the 30 test case configurations. EER for each test case are found at the intersections with the red line. Note that the best DET curve is very close to the bottom left corner, as happens when very few test vectors are misclassified; for this subject configuration, AUC was 0.9996 and EER was 0.0091. The worst case DET curve indicates that for an EER of 7% or less, the True Positive/Acceptance rate is at least 93%.

Of the methods explored here, GMM-UBM using a 2-channel combination yields the best results. A mean AUC of 0.9899 was found using GMM-UBM, which slightly outperforms 0.97 using SVMs [4]. Both studies used similar datasets: 8 subjects in our study vs 10 [4]; approximately 5 minutes of EEG recordings sampled at 512 Hz in our study vs approximately 2 minutes of EEG recordings sampled at 256 Hz [4]; lastly, windows of 2 s are used for feature extraction in our study vs windows of 1 s [4].

To provide a more comprehensive view of performance, AUC and EER are reported in terms of mean as well as 5th and 95th percentiles. For GMM-UBM the 5th percentile of AUC was higher than 0.97 indicating AUC to be highly concentrated near 1.

Note that the best 2-channel combination found by both the KNN and GMM-UBM approaches, based exclusively on analysis of the data, was Fc1-Pz. These electrodes are located between the Frontal and Central (between Frontal and Parietal) cortical regions, just left of center for Fc1, and over the center of the Parietal cortical region for Pz. Both the Frontal and Parietal regions are implicated in ADHD and involve brain networks and attention [16]. It is encouraging to find that the results of blind data analysis do not conflict with the neurobiology of ADHD.

V. CONCLUSION

Several classification methods were used to determine which selections of EEG channel combinations produce high accuracy of classification into ADHD and Non-ADHD subjects. For each of the methods investigated, accuracy was determined in terms of multiple measures of the distribution, for example the 5th and 95th percentiles and the mean, for the area under the curve (AUC) and equal error rate (EER). The KNN and GMM-UBM classifiers produced 2, 3, and 4 channel subset selections with quite a bit of overlap. Generally, the 2 and 3-channel selections appear to provide an excellent trade-off in terms of accurate performance and required computational effort. For the case of GMM-UBM, the pair Fc1-Pz outperformed all the other 2, 3, and 4-channel combinations. This combination of channels, along

with GMM-UBM, yielded a 5th percentile AUC value of 0.97, i.e. highly concentrated near 1. The 95th percentile EER was 0.0700. At an EER of 7% or less, the positive acceptance rate or detection probability for ADHD was at least 93%.

REFERENCES

- [1] *Diagnostic and statistical manual of mental disorders*. Washington, DC: American Psychiatric Association, 2013.
- [2] "Data and Statistics | ADHD | NCBDDD | CDC", *Cdc.gov*, 2016. [Online]. Available: <http://www.cdc.gov/ncbddd/adhd/data.html>. [Accessed: 10- Oct- 2016].
- [3] B. Abibullaev and J. An, "Decision Support Algorithm for Diagnosis of ADHD Using Electroencephalograms", *J Med Syst*, vol. 36, no. 4, pp. 2675-2688, 2011.
- [4] A. Mueller, G. Candrian, J. Kropotov, V. Ponomarev, and G. Baschera, "Classification of ADHD patients on the basis of independent ERP components using a machine learning system," *Nonlinear Biomedical Physics*, vol. 4, no. 1, p. S1, 2010.
- [5] K. Sadatnezhad, R. Boostani, and A. Ghanizadeh, "Classification of BMD and ADHD patients using their EEG signals," *Expert Systems with Applications*, vol. 38, no. 3, pp. 1956-1963, 2011.
- [6] J. Lopez Marcano, M. A. Bell, and A. A. (Louis) Beex, "Classification of ADHD and Non-ADHD using AR Models", *The Proceedings of Engineering in Medicine and Biology Conference (EMBC) 2016*. Pp. 363-366, Orlando, FL. Aug. 2016.
- [7] V. Monastra, J. Lubar, M. Linden, P. VanDeusen, G. Green, W. Wing, A. Phillips and T. Fenger, "Assessing attention deficit hyperactivity disorder via quantitative electroencephalography: An initial validation study," *Neuropsychology*, vol. 13, no. 3, pp. 424-433, 1999.
- [8] R. Chabot, H. Merkin, L. Wood, T. Davenport, and G. Serfontein, "Sensitivity and Specificity of QEEG in Children with Attention Deficit or Specific Developmental Learning Disorders," *Clinical EEG and Neuroscience*, vol. 27, no. 1, pp. 26-34, 1996.
- [9] S. K. Loo and M. Arns, Should the EEG-Based Theta to Beta Ratio Be Used to diagnose ADHD?, *The ADHD Report*, Vol. 23, No. 8, pp. 8-13, December 2015. doi: 10.1521/adhd.2015.23.8.8
- [10] S. Snyder, H. Quintana, S. Sexson, P. Knott, A. Haque, and D. Reynolds, "Blinded, multi-center validation of EEG and rating scales in identifying ADHD within a clinical sample," *Psychiatry Research*, vol. 159, no. 3, pp. 346-358, 2008. doi:10.1016/j.psychres.2007.05.006
- [11] M. I. Posner and S. E. Peterson, "The attention system of the human brain," *Annual Review of Neuroscience*, vol 13, pp. 25-42, 1990.
- [12] M. R. Rueda, J. Fan, B. D. McCandliss, J. Halparin,, D. Gruber, L.P. Lercari, and M. I. Posner, "Development of attentional networks during childhood," *Neuropsychologia*, vol. 42, pp. 1029-1040, 2004.
- [13] Reynolds, "Gaussian mixture models," in *Encyclopedia of Biometrics*, ed: Springer, 2009, pp. 659-663.
- [14] P. Raman, "Speaker Identification and Verification Using Line Spectral Frequencies", MS Thesis, Virginia Polytechnic Institute and State University, 2015.
- [15] P. M. Djuric and S. M. Kay, "Order selection of autoregressive models," *IEEE Transactions on signal processing* 40.11 (1992): 2829-2833.
- [16] ADHD Institute, <http://www.adhd-institute.com/burden-of-adhd/aetiology/neurobiology/>, [Accessed: 24- Oct- 2016].